

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 31.61 Seconds  
(without alignments)  
1633.955 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALCAALAG.....FRAGLRHGDEILAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802:\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2353	100.0	465	21	NGSP polypeptide o
2	2317	98.5	499	21	Neisseria gonorrhe
3	2254	95.8	499	21	Neisseria meningit
4	2248	95.5	499	21	Neisseria meningit
5	2244	95.4	499	21	Neisseria meningit
6	2234.5	95.0	498	21	Neisseria meningit
7	2202	93.6	499	21	Neisseria meningit
8	2197	93.4	499	21	Neisseria meningit
9	2130.5	90.5	475	21	Neisseria meningit
10	2092	88.9	414	21	NGSP polypeptide o
11	2002	85.1	448	21	Neisseria meningit

12	1832	77.9	370	21	AAV52996
13	962.5	40.9	460	20	AAV29294
14	731.5	31.1	503	17	AAV77434
15	705	30.0	474	22	AAV78605
16	685	29.1	491	20	AAV37309
17	633.5	26.9	475	12	AAV14625
18	632	26.9	463	17	AAV92690
19	629	26.7	437	21	AAV11834
20	627.5	26.7	455	22	AAV98946
21	624	26.5	437	21	AAV11837
22	624	26.5	437	21	AAV11838
23	623	26.5	437	21	AAV11836
24	622	26.4	437	21	AAV11835
25	603.5	25.6	491	19	AAV56771
26	573.5	24.4	443	20	AAV89850
27	565.5	24.0	451	20	AAV89977
28	559.5	23.8	387	19	AAV98445
29	542.5	23.1	433	21	AAV59353
30	542.5	23.1	433	21	AAV59355
31	533	22.7	433	21	AAV59354
32	453	19.3	549	22	AAV81139
33	449.5	19.1	397	21	AAV81772
34	448	19.0	320	21	AAV80962
35	448	19.0	362	21	AAV80958
36	448	19.0	394	21	AAV15193
37	448	19.0	451	21	AAV80957
38	448	19.0	480	18	AAV22849
39	448	19.0	480	21	AAV80956
40	448	19.0	480	21	AAV15191
41	447.5	19.0	355	22	AAV98947
42	445	18.9	480	21	AAV80964
43	443.5	18.8	405	21	AAV58844
44	442.5	18.8	413	22	AAV90724
45	442.5	18.8	441	22	AAV79050

## ALIGNMENTS

RESULT 1  
AAV83150  
ID AAV83150 standard; Protein; 465 AA.  
XX  
AC AAV83150;  
XX  
DT 24-JUL-2000 (first entry)  
XX  
DE NGSP polypeptide of Neisseria gonorrhoeae.  
XX  
KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
KW screening; probe; primer; prophylaxis; therapy.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN WO200012133-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US20070.  
XX  
PR 01-SEP-1998; 98US-0098685.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Jackson WJ, Harris AM;  
XX  
DR WPI: 2000-237782/20.  
XX  
N-PSDB; AA293414.  
PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
PT infections

Neisseria meningit  
Protein encoded by  
Heat shock protein  
Lawsonia intracell  
Protein involved i  
S. typhimurium Htr  
H. influenzae SB33  
Haemophilus influe  
E. coli growth and  
Haemophilus influe  
Haemophilus influe  
Haemophilus influe  
Escherichia coli s  
Antigen 1 from clu  
Protein encoded by  
H. pylori GHPO 536  
BAS011 protein se  
BAS011 protein se  
Mycobacterium tube  
Streptococcus pneu  
Human HtrA (high t  
Human HtrA (high t  
C-terminal domain  
Human HtrA (high t  
Osteoblast like ce  
Human HtrA (high t  
Human serine prote  
E. coli growth and  
Human HtrA (high t  
Breast and ovarian  
C glutamicum prote  
Corynebacterium gl

XX

Claim 5; Page 61-62; 68pp; English.

XX The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
 XX Arg-Gly-Asn motifs near the C-terminus which function as adherence  
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
 CC as a vaccine produces antibodies which inhibit binding of N.  
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or  
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments  
 CC can be used to immunise an animal and produce an immune response.  
 CC They can also be used as ligands to detect antibodies elicited in  
 CC response to Neisseria infections and also as antigens or immunogens  
 CC for inducing Neisseria-specific antibodies which are useful in  
 CC immunoassays to detect Neisseria in biological specimens. Nucleotides  
 CC encoding NGSP or its fragments can be used as probes to identify  
 CC Neisseria in biological specimens by hybridization or polymerase  
 CC chain reaction amplification. The NGSP polypeptide can also be used  
 CC in screening assays to identify agents and compounds which useful as  
 CC diagnostic, prophylactic or therapeutic agents against Neisseria  
 CC infection.

SQ Sequence 465 AA;

Query Match 100.0%; Score 2353; DB 21; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-190;  
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
 DQ 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
 QY 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120  
 DQ 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120  
 QY 121 GGLNFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSDVQSDVALLKIDA 180  
 DQ 121 GGLNFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSDVQSDVALLKIDA 180  
 QY 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
 DQ 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
 QY 241 INPGNSGGFLPNLKGQVINGINSQIYRSRGGFMGISFAIPDVMNVAEQLKNTGKVORGO 300  
 DQ 241 INPGNSGGFLPNLKGQVINGINSQIYRSRGGFMGISFAIPDVMNVAEQLKNTGKVORGO 300  
 301 LGVITQEVSYGLAQSGFLDKAGLAKILPGSPAERAGLQAGDIVLSLDGGETRSSGDL 360  
 DQ 301 LGVITQEVSYGLAQSGFLDKAGLAKILPGSPAERAGLQAGDIVLSLDGGETRSSGDL 360  
 QY 361 PYMVGAITPGKEVSLGVWRKGEEITIKALIGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
 DQ 361 PYMVGAITPGKEVSLGVWRKGEEITIKALIGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
 QY 421 AGITLQTHWDSGKHLVVVVSDDAABRAGLRHGDDEILAVRASPRQ 465  
 DQ 421 AGITLQTHWDSGKHLVVVVSDDAABRAGLRHGDDEILAVRASPRQ 465

RESULT 2

AAAY75748

ID AAAY75748 standard; Protein; 499 AA.

XX

AC AAAY75748;

XX

DT 21-MAR-2000 (first entry)

XX

DE Neisseria gonorrhoeae ORF 986 protein sequence SEQ ID NO:2968.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

KW antibacterial; gene therapy.

XX

OS Neisseria gonorrhoeae.

XX

PN WO957280-A2.

XX

PD 11-NOV-1999.

XX

PF 30-APR-1999; 99WO-US09346.

XX

PR 01-MAY-1998; 98US-0083758.

PR

PR 31-JUL-1998; 98US-0094869.

PR

PR 02-SEP-1998; 98US-0098994.

PR

PR 02-SEP-1998; 98US-0099062.

PR

PR 09-OCT-1998; 98US-0103749.

PR

PR 09-OCT-1998; 98US-0103794.

PR

PR 09-OCT-1998; 98US-0103796.

PR

PR 25-FEB-1999; 99US-0121528.

XX

(CHIR ) CHIRON CORP.

PA

(GENO-) INST GENOMIC RES.

XX

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;

PI

PI Petersen J, Pizza M, Rappuoli R, Rattl G, Scalato E, Scarselli M;

PI

PI Tettelin H, Venter JC;

XX

DR WPI: 2000-062150/05.

DR

DR N-PSDB; AAZ54510.

XX

XX Novel Neisserial polypeptides predicted to be useful antigens for

PT

PT vaccines and diagnostics -

XX

XX Claim 2; Page 1390; 1453pp; English.

PS

XX AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAZ54253 to AAZ55941

XX

XX represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention  
 CC may also be used in gene therapy protocols.

XX

SQ Sequence 499 AA;

Query Match 98.5%; Score 2317; DB 21; Length 499;  
 Best Local Similarity 99.1%; Pred. No. 1.7e-187;  
 Matches 459; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60  
 DQ 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGVSMLLPDFAQL 60

QY 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120  
 DQ 61 VQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSDPFYEFKRLVPMPEIQEEADD 120

QY 121 GGLNFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSDVQSDVALLKIDA 180  
 DQ 121 GGLNFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSDVQSDVALLKIDA 180

QY 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240  
 DQ 181 TEELPVVKIGNPKNLKPGEWAAIGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDDVA 240

QY 241 INPGNSGGFLPNLKGQVINGINSQIYRSRGGFMGISFAIPDVMNVAEQLKNTGKVORGO 300  
 DQ 241 INPGNSGGFLPNLKGQVINGINSQIYRSRGGFMGISFAIPDVMNVAEQLKNTGKVORGO 300

Db 241 inpgnsgplfnlkqgvvinsqlysrsgfmgisfaipidvamnvaqlkntgkvqrqg 300  
QY 301 LGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDITVLSDGGEIRSSGDL 360  
Db 301 lgvitqevsyglaqsfldkasgaliakilpgspaeraglaqgdvilsldggeirssgdl 360  
QY 361 PMVVGAIPTGKEVSLGVRKGEERTIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
Db 361 pvmvgaitpgkevsylgvrkgeertikaklgnaaehtgassktdapYTEQSGTFSVES 420  
QY 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRHGDEILAVRASP 463  
Db 421 agitlqhtdssgkhlvvvrsvdaaeraglrhgdeilavrgvqp 463

## RESULT 3

AAV52995

ID AAV52995 standard; Protein; 499 AA.

XX

AC AAV52995;

XX 21-FEB-2000 (first entry)

Neisseria meningitidis strain H44/76 BASB013 protein sequence.

XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial.

XX Neisseria meningitidis.

XX WO955872-A1.

XX 04-NOV-1999.

XX 20-APR-1999; 99WO-EP02765.

XX 23-APR-1999; 98GB-0008734.

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Ruelle J;

XX WPI; 2000-052809/04.

XX N-PSDB; AAZ33307.

XX Novel polynucleotides and polypeptides from Neisseria meningitis used  
XX to prepare vaccines against bacterial infections

XX Claim 3; Page 79-81; 94pp; English.

XX The present sequence represents a BASB013 polypeptide isolated from  
XX Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be  
XX employed as research reagents and material for the discovery of  
XX treatments and diagnostics for diseases, particularly human diseases.  
XX They can be used for diagnosis of disease, staging of disease, or  
XX determining response of an infectious organism to drugs. The  
XX polynucleotides may be used as a source for hybridisation probes, and  
XX for screening of genetic mutations, serotype, organism or strain  
XX identification, identification of mutation in BASB013 sequences, and as  
XX components of arrays which are useful for diagnostic and prognostic  
XX purposes. The polypeptides can be used to produce antibodies. The  
XX polypeptides can also be used in vaccine formulations, and to identify  
XX agonists and antagonists. The polypeptides, antibodies, agonists and  
XX antagonists (which are bacteriostatic) are used for the treatment and  
XX prevention of diseases such as upper respiratory tract infection,  
XX invasive bacterial diseases such as bacteraemia and meningitis, and for  
XX the development and screening of antibacterial drugs. They are also used  
XX in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
XX on in-dwelling devices, or to extracellular proteins on wounds, and to  
XX thus prevent tissue damage and/or block the normal progression of  
XX pathogenesis in infections initiated other than by the implantation of  
XX in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA;  
SQ

Query Match 95.8%; Score 2254; DB 21; Length 499;

Best Local Similarity 96.1%; Pred. No. 3.6e-182;

Matches 445; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 VFVKQYFALAALACALLAGCEKAGSFFGADKKEASVERIEHTKDDGGSVSMLLPDPFAQL 60

Db 1 mfkkygylalaalcaaslagcdkagsffgadkkeasfveriehtkddggsvsmllpdpfaql 60

QY 61 VQSEGPVVNIQAAPARTONGSNAETSDPLADSDPVEFFKRLVPMNPEIQEESADD 120

Db 61 vqsegpavvniqaapartqngsnaendsdpdiadndprieffkrlvpmnpeiqeeseadd 120

QY 121 GGLNFGSGPIISKNGYILNTHTVVGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180

Db 121 gglnfgsgfllskdgyilnthtvvtgmsikvllndkreytakligsdvqsdvallkida 180

QY 181 TEELPVVKIGNPNKLPGEWVAAGAPFGFDSVTAGIVSAKGRSLPNESYTPPIQTDVA 240

Db 181 teelpvkvkignpndlkpgewvaagapfgfdnsvtagivsakgrslpnesytpfiqtdva 240

QY 241 INPGNSGGPLFNKLGQVGVINSQIYSRSGFGMISFAIPIDVAMNVAEQLKNTGKVORQG 300

Db 241 inpgnsgplfnlkqgvvinsqlysrsgfmgisfaipidvammvaeqlkntgkvorgq 300

QY 301 LGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDITVLSDGGEIRSSGDL 360

Db 301 lgvitqevsyglaqsfldkagalliakilpgspaeraglaqgdvilsldggeirssgdl 360

QY 361 PMVVGAIPTGKEVSLGVRKGEERTIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420

Db 361 pvmvgaitpgkevsylgvrkgeertikaklgnaaehtgassktdapYTEQSGTFSVES 420

QY 421 AGITLQHTDSSGKHLVVVRVSDAAERAGLRHGDEILAVRASP 463

Db 421 agitlqhtdssgkhlvvvrsvdaaeraglrhgdeilavrgvqp 463

## RESULT 4

AAV5749

ID AAV5749 standard; Protein; 499 AA.

XX

AC AAV5749;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2970.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099082.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 25-FEB-1999; 98US-0103796.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.



QY 181 TEELPVVKIGNPKNLKPGWAAIGAPFGFNSVTAGIVSAKGRSLPNESYTPFIOTDVA 240  
|||||  
Db 181 teelpvkvignpdklpewvaalgapfgfndsvtagvksakgrslpnesytpfiqtdva 240  
QY 241 INPNSGGPLNKLKQGVVINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTKGVORGO 300  
|||||  
Db 241 inpgnsggplfnlkgvvgvlnsqiysrsgfmgisfaipidvamnvaeklntgkvqrgq 300  
QY 301 LGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
|||||  
Db 301 lgviiqevsyglagsfgldkaggaliaakilpgspaeragragdivlsldggeirssgd 360  
QY 361 PMVVGAIPTPKVESLGVWRKGEETIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
|||||  
Db 361 pvmvgaitpgkveslgvwrkgeetikvklgnaeahigassktdapYTEQSGTFSVES 420  
QY 421 AGITLQHTDSSGKHLVVVRVSDAERAGRLRHGDEILAVRASP 463  
|||||  
Db 421 agitlqhtdssgghlvvrvrsvdaaeragrrgdellavrgvp 463

RESULT 6  
AAY70413  
ID AAY70413 standard; Protein; 498 AA.  
AC AAY70413;  
XX  
DT 03-JUL-2000 (first entry)  
XX  
DE Neisseria meningitidis NMASP protein-2.  
XX  
KW NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.  
XX  
OS Neisseria meningitidis.  
XX

\*PH Key Location/Qualifiers  
FT Misc-difference 8 /note= "Encoded by TTC"  
FT Misc-difference 17 /note= "Encoded by TTG"  
FT Misc-difference 22 /note= "Encoded by GAA"  
FT Misc-difference 42 /note= "Encoded by GAA"  
FT Misc-difference 58 /note= "Encoded by GCC"  
FT Misc-difference 84 /note= "Encoded by GGC"  
FT Misc-difference 134 /note= "Encoded by AAC"  
FT Misc-difference 145 /note= "Encoded by GCC"  
FT Misc-difference 194 /note= "Encoded by AAT"  
FT Misc-difference 218 /note= "Encoded by GGCATC"  
FT Misc-difference 321 /note= "Encoded by AGC"  
FT Misc-difference 340 /note= "Encoded by AGG"  
FT Misc-difference 387 /note= "Encoded by GCC"  
FT Misc-difference 396 /note= "Encoded by ACC"  
FT Misc-difference 433 /note= "Encoded by AAA"  
FT Misc-difference 451 /note= "Encoded by CAC"  
FT Misc-difference 459 /note= "Encoded by AGG"

Misc-difference 460 /note= "Encoded by GCA"  
FT Misc-difference 461 /note= "Encoded by AGT"  
FT Misc-difference 463 /note= "Encoded by CGT"  
FT Misc-difference 464 /note= "Encoded by CAA"  
FT Misc-difference 465.498  
FT /note= "Encoded by AGGCAAGTCCCGCTCAA"  
XX WO200012535-A2.  
XX 09-MAR-2000.  
PD  
XX  
PF 01-SEP-1999; 99WO-US19663.  
XX  
PR 01-SEP-1999; 98US-0098685.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Jackson WJ, Harris AM;  
XX  
DR WPI: 2000-256581/22.  
DR N-PSDB; AAZ51538.  
XX  
FT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and  
FT antibodies, useful in vaccines against infection  
XX  
PS Claim 5; Page 70-71; 75pp; English.  
XX  
CC The present sequence is the Neisseria meningitidis NMASP protein.  
CC NMASP is a non-cytosolic protein, with antibacterial and  
CC antiinflammatory activity. It shows sequence similarity to E. coli  
CC Degp (Htra) protein. NMASP proteins can be used as ligands to detect  
CC antibodies elicited in response to N. meningitidis infections. Cytotoxic  
CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA  
CC may be used for diagnosis, therapy or prophylaxis of Neisserial  
CC infections such as, bacterial meningitidis and septicaemia.  
XX  
SQ Sequence 498 AA;

Query Match 95.0%; Score 2234.5; DB 21; Length 498;  
Best Local Similarity 95.9%; Pred. No. 1.6e-180;  
Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1;

QY 1 VFKKYQYFALAAALCAGCEKAGSPFGADKKEASFVERIEHTWKDDGVSVMLLPDFAQL 60  
|||||  
Db 1 vfkkyyalalaalcaaslagcdkagsffgdkkkaasfverikhtkddgsvmllpdfvql 60  
QY 61 VQSEGPVAVNIOAAPARTONGSGNAETDSDPLADSDPFYEFFKRLVPMPEIQEEADD 120  
|||||  
Db 61 vqsegpavvniqaapartngsgsnaetdsdpladsdpfyeffkrlvpmpeiqeeadd 120  
QY 121 GGLNFGSGFTIISKNGYILTNTHVVAGMSIKVLNDRKREYTAKLIGSDVQSDVALIKIDA 180  
|||||  
Db 121 gglnfgsgftiiskngdyilltnthvvtgmgsikvlndrkreytakligsdvqsdvalikida 180  
QY 181 TEELPVVKIGNPKNLKPGWAAIGAPFGFNSVTAGIVSAKGRSLPNESYTPFIOTDVA 240  
|||||  
Db 181 teelpvkvignpdklpewvaalgapfgfndsvtagvksakgrslpnesytpfiqtdva 240  
QY 241 INPNSGGPLNKLKQGVVINSQIYSRSGFMGISFAIPIDVAMNVAEOLKNTKGVORGO 300  
|||||  
Db 241 inpgnsggplfnlkgvvgvlnsqiysrsgfmgisfaipidvamnvaeklntgkvqrgq 299  
QY 301 LGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
|||||  
Db 301 lgviiqevsyglagsfgldkaggaliaakilpgspaeragragdivlsldggeirssgd 359  
QY 361 PMVVGAIPTPKVESLGVWRKGEETIKAKLGNAAEHTGASSKTDAPYTEQSGTFSVES 420  
|||||

Db 360 pvmvgaitpgkevalgwwrkgeetkikvklgnaaehigassktddeapyeqsgtgsfsvs 419

QY 421 AGITLOTHTDSSGRHLVVVYRSDAAERAGLRHGDEILAVRASP 463

Db 420 agitlqhtdssghlvvvrvsdaaeraglrhgdeilavgvvp 462

RESULT 7

ID AAY52994 standard; Protein: 499 AA.

XX AC AAY52994;

XX DT 21-FEB-2000 (first entry)

XX DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.

XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;

XX OS antibiotic; upper respiratory tract infection; bacteraemia; meningitis;

XX PN invasive bacterial disease; antibacterial.

XX OS Neisseria meningitidis.

XX PN WO9955872-A1.

XX PD 04-NOV-1999.

XX PF 20-APR-1999; 99WO-EP02765.

XX PR 23-APR-1998; 98GB-0008734.

XX \*A (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX DR WPI: 2000-052809/04.

XX DR N-PSDB; AA233306.

XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used

XX PT to prepare vaccines against bacterial infections -

XX PS Claim 3; Page 77-78; 94pp; English.

XX CC The present sequence represents a BASB013 polypeptide isolated from

XX CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be

XX CC employed as research reagents and material for the discovery of

XX CC treatments and diagnostics for diseases, particularly human diseases.

XX CC They can be used for diagnosis of disease, staging of disease, or

XX CC determining response of an infectious organism to drugs. The

XX CC polynucleotides may be used as a source for hybridisation probes, and

XX CC for screening of genetic mutations, serotype, organism or strain

XX CC identification, identification of mutation in BASB013 sequences, and as

XX CC components of arrays which are useful for diagnostic and prognostic

XX CC purposes. The polypeptides can be used to produce antibodies. The

XX CC polypeptides can also be used in vaccine formulations, and to identify

XX CC agonists and antagonists. The polypeptides, antibodies, agonists and

XX CC antagonists (which are bacteriostatic) are used for the treatment and

XX CC prevention of diseases such as upper respiratory tract infection,

XX CC invasive bacterial diseases such as bacteraemia and meningitis, and for

XX CC the development and screening of antibacterial drugs. They are also used

XX CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins

XX CC on in-dwelling devices, or to extracellular proteins on wounds, and to

XX CC thus prevent tissue damage and/or block the normal progression of

XX CC pathogenesis in infections initiated other than by the implantation of

XX CC in-dwelling devices or by other surgical techniques.

XX SQ Sequence 499 AA;

Query Match 93.6%; Score 2202; DB 21; Length 499;

Best Local Similarity 94.0%; Pred. No. 9.1e-178;

Matches 435; Conservative 10; Mismatches 18; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAAALCAALLACGECAGSGFFGADKKKEASFVERIEHTKDDGVSMLLPDFAQL 60

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

1 mfkkyqylalaaicaaslagckagsffgadkkaasfverlehckddgsvsmllpdrvql 60

QY 61 VQSEGAUVVNIQAAPARTQNGSCNAETDSDPLADSDPFYEFFKRLVPMPEIQOEADD 120

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

61 vqsegpavvniqaapartqngsgnaetdspdpladsdpfyeffkrlvpmpeiqpgeadd 120

QY 121 GGLNFGSGFIISKNGYILTNTHVVAGMSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

121 gglngsgfiiiskdgyiltntthvvtgmgsikvllndkreytakligsdvqsdvallkida 180

QY 181 TEELPVVYKIGNPNKLPGEWVAAGAPFGFDSNVSTAGIVSAKGRSLPNESYTPFTQDVA 240

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

181 teelpvvkignpkdlkpgewvaalgapfgfndsvtagivsaakgrslpnesytpfqtqtdva 240

QY 241 INPNSGGPLFLNLKGOVVGINSIQIYSRSGFGMISFAIPIDVAMNVAEOLKNTCKVORGO 300

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

241 inpgnsggplfnlkqvgvinsqlyrsrgfmgisfalpdivamnvaeqlkntgkvrgq 300

QY 301 LGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

301 lgviiqevsyglaqsgfldkagallakilpgspaeraglgagdivisldggeirssgdl 360

QY 361 PVMVGAITPGKEVSLGVMRKGEETIKAKLGNAEHTGASKTDEAPYTEQSGTFSVES 420

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

361 pvmvgaitpgkevsylgvmrkgeetivaklgnaehtgaseqtdgsktdeapyaehqsgtfsves 420

QY 421 AGITLOTHTDSSCKHLVVVVRVSDAAERAGLRHGDEILAVRASP 463

Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

421 agitlothtdssckhlvvvrvsdaaeraglrhgdeilavrasp 463

RESULT 8

AAY52993

ID AAY52993 standard; Protein: 499 AA.

XX AC AAY52993;

XX DT 21-FEB-2000 (first entry)

XX DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.

XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;

XX OS antibiotic; upper respiratory tract infection; bacteraemia; meningitis;

XX PN invasive bacterial disease; antibacterial.

XX OS Neisseria meningitidis.

XX PN WO9955872-A1.

XX PD 04-NOV-1999.

XX PF 20-APR-1999; 99WO-EP02765.

XX PR 23-APR-1998; 98GB-0008734.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Ruelle J;

XX DR WPI: 2000-052809/04.

XX DR N-PSDB; AA233306.

XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used

XX PT to prepare vaccines against bacterial infections -

XX PS Claim 5; Page 74-76; 94pp; English.

XX CC The present sequence represents a BASB013 polypeptide isolated from

XX CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be

XX CC employed as research reagents and material for the discovery of

XX CC treatments and diagnostics for diseases, particularly human diseases.

They can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB013 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases such as upper respiratory tract infection, invasive bacterial diseases such as bacteraemia and meningitis, and for the development and screening of antibacterial drugs. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

```

Query Match          93.4%;   Score 2197;   DB 21;   Length 499;
Best Local Similarity 93.7%;   Pred. No. 2.4e-177;
Matches 434;   Conservative 10;   Mismatches 19;   Indels 0;   Gaps 0;

1  VFKKYQYFALAAALCAGCEKAGCSFTGCADKKEASFVERIEHTTKDDGSVSMLLPDPFAQL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1  mfkkyqylalaaacaaslagcdkaqsfgyadkkaasfveriehtkddgsvsmllpdfvql 60

61  VOSEGPVAVVNTQAAPAPRTQTGNSGNAETSDPLADSDPFYEFFKRLVPMNPEIPOEEADD 120
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61  vqsegpavvnvigaapaprtqtgnsгнаetdsdpadsdpfyefkrlvpmnpeipqeeadd 120

121  GGLNFGSGFIISKNGYILTNTHVAVAGMSGIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
121  gglngfsgfiiiskdgyiltnthvvtgmsgikvllndkreytakligsdvsdvalkkida 180

181  TEELPVVKIGNPNKLNKPGEWAAIAGAPFGFDSNVTAGIVSAKGRSLPNESYTPFIQTDA 240
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
181  teelpvvkignpnkdlkpgewaaigapfgfdnsvtagivsakgrslpnesytpfiqtava 240

241  INPNGSGGPLFNLKGVGVGINSQIYVSRSGFGMISFAIPIDVAMVAEQLNKTGVQRGQ 300
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
241  inpngsggplfnlkgvvgvinsqilysrsggfmgisfaipidvamvnaeqlnktgkvqrgq 300

301  LGVITQEVSVYGLAQSGFLDKAGSALIATILPGSPAERAGLAGQDITVLSLDGGEIRSSGDL 360
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
301  lgvitqevsyglqsgfgldkagaliatilpgspaeraglagdvlslsdggeirssgdl 360

361  PVMVCAITPGKEVSLGVWRKKGEEITIKAKLGNAAEHTCASSKTDRAPTYEQSGSGFVS 420
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
361  pvmvcaitpgkevsilgvwrkgkeltvavklnaaseqtssssepdhapyaehtgsgfvs 420

421  AGITLOTHTDSSGKHLVVVRVSDAAERAGLRHGDEILAVRAPR 463
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
421  agitlothtdssgqrlvvvrvsqaaeraglrhgdeilavqgvp 463

```

RESULT	9
AA170414	
ID	AA170414 standard; Protein; 475 AA.
XX	
XX	AA170414;
XX	
DT	03-JUL-2000 (first entry)
XX	
DE	Neisseria meningitidis NNA5P protein

XX NNASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NNASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.  
XX

OS	Neisseria meningitidis.
XX	
Key	Location/Qualifiers
FH	Misc-difference 19
FT	/note= "Encoded by GAA"
FT	Misc-difference 35
FT	/note= "Encoded by GCC"
FT	Misc-difference 61
FT	/note= "Encoded by GGC"
FT	Misc-difference 111
FT	/note= "Encoded by AAC"
FT	Misc-difference 122
FT	/note= "Encoded by GCC"
FT	Misc-difference 171
FT	/note= "Encoded by AAT"
FT	Misc-difference 194
FT	/note= "Encoded by GGCATC"
FT	Misc-difference 298
FT	/note= "Encoded by AGC"
FT	Misc-difference 317
FT	/note= "Encoded by AGG"
FT	Misc-difference 364
FT	/note= "Encoded by GCC"
FT	Misc-difference 373
FT	/note= "Encoded by ACC"
FT	Misc-difference 410
FT	/note= "Encoded by AAA"
FT	Misc-difference 428
FT	/note= "Encoded by CAC"
FT	Misc-difference 436
FT	/note= "Encoded by AGG"
FT	Misc-difference 437
FT	/note= "Encoded by GCA"
FT	Misc-difference 438
FT	/note= "Encoded by AGT"
FT	Misc-difference 440..475
FT	/note= "Encoded by CGTCAA"
XX	
PN	WO200012535-A2.
XX	
PD	09-MAR-2000.
XX	
PF	01-SEP-1999; 99WO-US19663.
XX	
PR	01-SEP-1998; 98US-0098685.
XX	(ANTE-) ANTEX BIOLOGICS INC.
PA	Jackson WJ, Harris AM;
PI	
XX	WPI; 2000-256581/22.
DR	N-PSDB; AAZ51539.
XX	
PT	Neisseria meningitidis NMASP polypeptide,
PT	antibodies, useful in vaccines against in
XX	
PS	Claim 5; Page 71-73; 75pp; English.
XX	
CC	The present sequence is the Neisseria men
CC	NMASP is a non-cytosolic protein, with an
CC	antiinflammatory activity. It shows sequen
CC	DegP (HtrA) protein. NMASP proteins can b
CC	antibodies elicited in response to N. men
CC	anti-NMASP antibodies can be used as vacc
CC	may be used for diagnosis, therapy or prop
CC	infections such as, bacterial meningitidi
XX	
SO	Sequence 475 AA:

Query Match 90.5%; Score 2130.5; DB 21; Length 475;  
Best Local Similarity 96.4%; Pred. NO. 9.6e-172;  
Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 24 AGSFFGADKKEASFVERIEHTKDDGSVMLLPDPFAQLVQSEGPVAVNIQAAPAPRTQMS 83  
 DDb 1 agsffgadkkesfveriehtkddgsvmllpdfvqlvqsegpavvniqaapaprtqms 60  
 QY 84 GNAETSDPLADSDPYEFKRLVPMPEIQBEADGGLNFGSGFTISKNGYILTNTHV 143  
 DDb 61 snaetdspladsdpyeffkrlvpmpeipqeeaddgglnfgsgftiskngyiltnthv 120  
 QY 144 VAGMGKIVLLNDKREYAKLIGSDVQSDVALLKIDATEELPVVKIGNPKLKPGEWAA 203  
 DDb 121 vtmgskivllndkreytaklignsdvqsdvallkideateelpvkvignpkdkpgewaa 180  
 QY 204 IGAPFGFDSVTVAGIVSAKGRSLPNESYTPFIQTDVAINPNSGGPLFNKLGQVVGINSQ 263  
 DDb 181 igapfgfdsvtvtagivsaakgrslpnesytpfiqtdvainpnsggplfnlkgqvvginsq 239  
 QY 264 IYSRSGFGWGISFAIPIDVAMNVAELKNTKQVQSGGLVQIIEVSYGLAQSGFLDKASG 323  
 DDb 240 iysrsgfgwgisfaipidvamnvaeklntkvqrgqlvqievsyglqsgfldkag 299  
 QY 324 ALIAKILPGSPARAGLQAGDIVLSLDGGGIRSGGDLPMVVGAIPTGKEVSLGVWRKGE 383  
 DDb 300 aliakilpgspaeragragdivlsldgggirsggdlpmvvgaitpgkevsldgvwrkgee 359  
 QY 384 ITKAKLGNAAHTGASSTDEAPYTEQSGTFSVESAGITLQHTDSSGKHLVVRVSD 443  
 DDb 360 itikvklgnaahtgassktdeapylteqsgtfsvesagitlqhtdssgkhlvvrvsd 419  
 QY 444 AAERAGLRHGDDELAVRASPRQ 463  
 DDb 420 aaeraglrhgdellavrgqp 439

## RESULT 10

AA83151  
 ID AAY83151 standard; Protein; 414 AA.

XX AC AAY83151;

XX 24-JUL-2000 (first entry)

XX NGSP polypeptide of Neisseria gonorrhoeae.

XX NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
 KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
 screening; probe; primer; prophylaxis; therapy.

XX Neisseria gonorrhoeae.

XX WO200012133-A1.

PN 09-MAR-2000.

XX 01-SEP-1999; 99WO-US20070.

XX 01-SEP-1998; 98US-0098685.

XX (ANTE-) ANTEX BIOLOGICS INC.

XX Jackson WJ, Harris AM;

XX WPI; 2000-237782/20.

XX N-PSDB; AA293415.

XX Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
 PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
 PT infections

XX Claim 5; Page 63-64; 68pp; English.

XX The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
 CC Arg-Gly-Asp motifs near the C-terminus which function as adherence

CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
 CC as a vaccine produces antibodies which inhibit binding of N.  
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or  
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments  
 CC can be used to immunise an animal and produce an immune response.  
 CC They can also be used as ligands to detect antibodies elicited in  
 CC response to Neisseria infections and also as antigens or immunogens  
 CC for inducing Neisseria-specific antibodies which are useful in  
 CC immunassays to detect Neisseria in biological specimens. Nucleotides  
 CC encoding NGSP or its fragments can be used as probes to identify  
 CC Neisseria in biological specimens by hybridization or polymerase  
 CC chain reaction amplification. The NGSP polypeptide can also be used  
 CC in screening assays to identify agents and compounds which useful as  
 CC diagnostic, prophylactic or therapeutic agents against Neisseria  
 CC infection.

XX Sequence 414 AA:

Query Match 88.9%; Score 2092; DB 21; Length 414;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-168;  
 Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 MLLPFAQLVQSEGPVAVNIQAAPAPRTQNGSGNAETSDPLADSDPYEFKRLVPMNP 111  
 DDb 1 mllpfaqlvqsegpavvniqaapaprtqngsgnaetdsdpladsdpyeffkrlvpmnp 60  
 QY 112 EIPOEADDGGLNFGSGFTISKNGYILTNTHVVGMSIKVLLNDKREYAKLIGSDVQS 171  
 DDb 61 elpoeaddgglnfgsgftiskngyiltnthvvagmsikvllndkreycakligsdvqs 120  
 QY 172 DVALLKIDATEELPVVKIGNPKLKPGEWAAIGAPFGFDSVTVAGIVSAKGRSLPNESY 231  
 DDb 121 dvallkideateelpvkvignpkpgewvaagapfgfdsvtvtagivsaakgrslpnesy 180  
 QY 232 TPFITQTDVAINPNSGGPLFNKLGQVVGINSQIYSRSGFGMGSFAIPIDVAMNVAEQLK 291  
 DDb 181 tpfitqtdvainpnsggplfnlkgqvvginsqlysrsgfgmgsfaipidvamnvaekl 240  
 QY 292 NTGKVQVQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDG 351  
 DDb 241 ntgkvqrgqlgvilqevsyglqsgfldkasgaliakilpgspaeraglgagdivlsldg 300  
 QY 352 GEIRSGGDLPMVVGAIPTGKEVSLGVWRKGEIITKAKLGNAAHTGASSTDEAPYTEQ 411  
 DDb 301 geirsggdlpmvvgaitpgkevsldgvwrkgeelitkaklgnaahtgassktdeapylteq 360  
 QY 412 QSGTFSVESAGITLQHTDSSGKHLVVRVSDAAERAGLRHGDDELAVRASPRQ 465  
 DDb 361 qsgtfsvesagitlqhtdssgkhlvvrvsdaaeraglrhgdellavrasprq 414

## RESULT 11

AA870409

ID AAY70409 standard; Protein; 448 AA.

XX AC AAY70409;

XX 03-JUL-2000 (first entry)

XX Neisseria meningitidis NMASP protein-1.

XX NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
 KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
 KW Neisserial infection; meningitidis; septicaemia.

XX Neisseria meningitidis.

XX Key . Location/Qualifiers

FT Misc-difference 167

XX /note= "Encoded by NTC"

PN WO200012535-A2.



```
XX PD 09-MAR-2000.
XX PF 01-SEP-1999; 99WO-US19663.
XX PR 01-SEP-1998; 98US-0098685.
XX PA (ANTE-) ANTEX BIOLOGICS INC.
XX PI Jackson WJ, Harris AM;
XX DR WPI: 2000-256581/22.
XX DR N-PSDB: AAZ51533.
XX PT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
XX PT antibodies, useful in vaccines against infection
XX PS Claim 5; Page-; 75pp; English.
XX CC The present sequence is the Neisseria meningitidis NMASP protein.
XX CC NMASP is a non-cytosolic protein, with antibacterial and
XX CC anti-inflammatory activity. It shows sequence similarity to E. coli
XX CC DegP (HtrA) protein. NMASP proteins can be used as ligands to detect
XX CC antibodies elicited in response to N. meningitidis infections. Cytotoxic
XX CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
XX CC may be used for diagnosis, therapy or prophylaxis of Neisserial
XX CC infections such as, bacterial meningitis and septicemia.
XX CC Note: The protein represented in SEQ ID NO:2 of the specification is
XX CC erroneous. The present sequence is the decoded version of the nucleotide
XX CC represented in AAZ51533.
XX CC Sequence 448 AA;
XX SQ
Query Match 85.1%; Score 2002; DB 21; Length 448;
Best Local Similarity 96.4%; Pred. No. 6.6e-161;
Matches 397; Conservative 3; Mismatches 12; Indels 0; Gaps 0;
Qy 52 MLLPFAQLVQSEGPVAVNIQAAPRTONGSGNAETSDPLADSPFFVEFFKRLVPNMP 111
Db 1 mllpfdvqlvqsegpavvniqaaprtongsgnaetdsdpladsdpffveffkrlvpnmp 60
Qy 112 EIPOEADDGGLNFCGFTISKNGYILTNTHVVAGMGSIKVLLNDKREYTKALIGSDVQS 171
Db 61 elpoeaddgglnfcgftiskngyiltnthvvtgmsikvllndkretytkaligsdvqs 120
Qy 172 DVALLKIDATEELPVYKIGNKPNLKPGEWVAATGAPFGFDSNVTAGIVSAKGRSLPNESY 231
Db 121 dvallkideateelpvkkignpkdlkpgewvaatgafgfdnsvtagvsakgrslpnesy 180
Qy 232 TPFTOTDVAINPGNSGGPLFNLKGQVVGINSQIYSRSGFGMGLSFAIPIDVAMNVAEQLK 291
Db 181 tpfqtvdvainpgnsggplfnlkgqvvginsqiyysrgfgmglisfaipidvamnvaekl 240
Qy 292 NTGKVQRGOLGVIQEVSVGLAQSGELDKASGALITAKILPGSPAERAGLQAGDIVLSLDG 351
Db 241 ntgkvqrqglvqldevsyglqsfldkagallakilpgspaeragragdivlsldg 300
Qy 352 GEIRSSGDLPMVMGATITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSKTEAPYTEQ 411
Db 301 geirssgdlpmvmgatitpgkevsylvwrkgeeitikvklgnaaehlgasskdeapYTEQ 360
Qy 412 QSGTFSVESAGITLTQHTDSSGKHLVWVRVSDAAERAGLRHGDIELAVRASP 463
Db 361 qsgtfsvesagitltqhtdssghlvwvrsvsdAAERaglrghgdellavqgvp 412
RESULT 12
AA52996
ID AA52996 standard; Protein: 370 AA.
XX AC
XX AA52996;
```

```
DT '21-FEB-2000 (first entry)
XX Neisseria meningitidis BASB013-C protein sequence.
XX DE
XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
XX KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
XX KW invasive bacterial disease; antibacterial.
XX OS Neisseria meningitidis.
XX PN WO9955872-A1.
XX PD 04-NOV-1999.
XX PF 20-APR-1999; 99WO-EP02765.
XX PR 23-APR-1998; 98GB-0008734.
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX PI Ruelle J;
XX WPI: 2000-052809/04.
XX DR N-PSDB: AAZ33308.
XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used
XX PT to prepare vaccines against bacterial infections
XX PS Example 2: Page 82-83; 94pp; English.
XX CC
XX CC The present sequence is a conserved BAS013-C polypeptide isolated
XX CC from Neisseria meningitidis. BAS013 polynucleotides and polypeptides
XX CC may be employed as research reagents and material for the discovery of
XX CC treatments and diagnostics for diseases, particularly human diseases.
XX CC They can be used for diagnosis of disease, staging of disease, or
XX CC determining response of an infectious organism to drugs. The
XX CC polynucleotides may be used as a source for hybridisation probes, and
XX CC for screening of genetic mutations, serotype, organism or strain
XX CC identification, identification of mutation in BAS013 sequences, and as
XX CC components of arrays which are useful for diagnostic and prognostic
XX CC purposes. The polypeptides can be used to produce antibodies. The
XX CC polypeptides can also be used in vaccine formulations, and to identify
XX CC agonists and antagonists. The polypeptides, antibodies, agonists and
XX CC antagonists (which are bacteriostatic) are used for the treatment and
XX CC prevention of diseases such as upper respiratory tract infection,
XX CC invasive bacterial diseases such as bacteraemia and meningitis, and for
XX CC the development and screening of bacteria to eukaryotic matrix proteins
XX CC on in-dwelling devices, or to extracellular proteins on wounds, and to
XX CC thus prevent tissue damage and/or block the normal progression of
XX CC pathogenesis in infections initiated other than by the implantation of
XX CC in-dwelling devices or by other surgical techniques.
XX SQ Sequence 370 AA;
Query Match 77.9%; Score 1832; DB 21; Length 370;
Best Local Similarity 97.0%; Pred. No. 1.2e-146;
Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
Qy 1 VFKEYQYFALAAALCAGCEKAGSFFGADKKKEASFVERIETHTKDDGVSMLLPDFAQL 60
Db 1 mfkkyqylalaalcaaslagcdkagsffgadkkesafveriehtkddgvsmlldpfaql 60
Qy 61 VQSGPAAVNIQAAPRTONGSGNAETSDPLADSPFFVEFFKRLVPNMPPEQEAADD 120
Db 61 vqsgpavvniqaaprtongsgnaetdsdpladsdpffveffkrlvpmppegeadd 120
Qy 121 GGLNFGSGFTISKNGYILTNTHVVAGMGSIKVLLNDKREYTKALIGSDVQSVALLKIDA 180
Db 121 gglngsgftiskngyiltnthvvtgmsikvllndkretytkaligsdvqsvallkida 180
Qy 181 TEELPVYKIGNKPNLKPGEWVAATGAPFGFDSNVTAGIVSAKGRSLPNESYTPFIOTDVA 240
```





---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 16.36 Seconds  
(without alignments)  
694,248 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQVFALALCAALLAG.....ERAGLRHGEDEILAVRASPRQ 465

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pap.\*
  - 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pap.\*
  - 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pap.\*
  - 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pap.\*
  - 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pap.\*
  - 6: /cgn2\_6/ptodata/2/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	990.5	42.1	460	4	US-09-199-637A-132
2	731.5	31.1	503	1	US-08-245-294-8
3	731.5	31.1	503	1	US-08-474-499-8
4	731.5	31.1	503	1	US-08-307-279A-8
5	731.5	31.1	503	5	PCT-US95-06211-8
6	646	27.5	463	1	US-08-485-569-2
7	646	27.5	463	1	US-08-480-993-2
8	646	27.5	463	2	US-07-903-079B-2
9	633.5	26.9	475	1	US-08-350-741-2
10	633.5	26.9	475	2	US-08-463-875A-2
11	632	26.9	463	1	US-08-278-091-2
12	632	26.9	463	1	US-08-483-859-2
13	632	26.9	463	1	US-08-472-173-2
14	632	26.9	463	2	US-08-487-167-2
15	632	26.9	463	2	US-08-482-816-2
16	632	26.9	463	2	US-08-296-149-2
17	632	26.9	463	2	US-08-801-499-2
18	632	26.9	463	2	US-08-615-271-2
19	632	26.9	463	3	US-09-074-660-2
20	632	26.9	463	3	US-09-106-468-2
21	632	26.9	463	4	US-09-106-466A-2
22	632	26.9	463	4	US-09-106-467-2
23	632	26.9	463	4	US-09-106-468-2
24	631.5	26.8	475	1	US-08-278-091-6
25	631.5	26.8	475	1	US-08-483-859-6
26	631.5	26.8	475	1	US-08-472-173-6
27	631.5	26.8	475	2	US-08-487-167-6

28	631.5	26.8	475	2	US-08-482-816-6	Sequence 6, Appli
29	631.5	26.8	475	2	US-08-296-149-6	Sequence 6, Appli
30	631.5	26.8	475	2	US-08-801-499-6	Sequence 6, Appli
31	631.5	26.8	475	2	US-08-615-271-6	Sequence 6, Appli
32	631.5	26.8	475	3	US-09-074-660-6	Sequence 6, Appli
33	631.5	26.8	475	3	US-09-074-659-6	Sequence 6, Appli
34	631.5	26.8	475	3	US-09-106-468-6	Sequence 6, Appli
35	631.5	26.8	475	4	US-09-106-466A-6	Sequence 6, Appli
36	631.5	26.8	475	4	US-09-106-467-6	Sequence 6, Appli
37	606.5	25.8	472	1	US-08-278-091-5	Sequence 5, Appli
38	606.5	25.8	472	1	US-08-483-859-5	Sequence 5, Appli
39	606.5	25.8	472	1	US-08-472-173-5	Sequence 5, Appli
40	606.5	25.8	472	2	US-08-487-167-5	Sequence 5, Appli
41	606.5	25.8	472	2	US-08-482-816-5	Sequence 5, Appli
42	606.5	25.8	472	2	US-08-296-149-5	Sequence 5, Appli
43	606.5	25.8	472	2	US-08-801-499-5	Sequence 5, Appli
44	606.5	25.8	472	2	US-08-615-271-5	Sequence 5, Appli
45	606.5	25.8	472	3	US-09-074-660-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-09-199-637A-132  
; Sequence 132, Application US/09199637A  
; Patent No. 6355411  
; GENERAL INFORMATION:  
; APPLICANT: Ausubel, Frederick  
; APPLICANT: Goodman, Howard M.  
; APPLICANT: Rahme, Laurence G.  
; APPLICANT: Mahajan-Miklos, Shalina  
; APPLICANT: Tan, Man-Wah  
; APPLICANT: Cao, Hul  
; APPLICANT: Drenkard, Eliana  
; APPLICANT: Tsongalis, John  
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID  
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF  
; FILE REFERENCE: 00786/361002  
; CURRENT APPLICATION NUMBER: US/09/199,637A  
; CURRENT FILING DATE: 1998-11-25  
; PRIOR APPLICATION NUMBER: 60/066,517  
; PRIOR FILING DATE: 1997-11-25  
; NUMBER OF SEQ ID NOS: 437  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 132  
; LENGTH: 460  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-199-637A-132

Query Match	42.1%	Score	990.5	DB	4	Length	460
Best Local Similarity	52.5%	Pred. No.	2.6e-84				
Matches	220	Conservative	55	Mismatches	111	Indels	33
Gaps	11						
Qy	54	LPDFAQLVQSEGPAYVNI	---QAAPAPRTONGSGNAETSDPLADSDP	-FYEFKRLVNP	109		
Db	27	LPDFTPLVEQASPAVVNISTROKLP	-----DRAMARQLSIPDLEGLPPMFRDLERTIPQ	82			
Qy	110	MPEIP	---QEADDCGLNFGSFIISKNGYI	LTNTHVAGMGSIKVLLNDKREYTA	166		
Db	83	VPRNPGQOREAQ	-----SLGSGFIISNDGYI	LTNNHVADAEILYRLSDRSEHKAKLVG	138		
Qy	167	SDVQSDVALLKIDAPEELPVVYKGNPKNLKPGEWAAICAPFGFNSVTAGIVSAKGRSL	226				
Db	139	ADPRSDVAVLKIEA	-KNLPTLKLGSNKLKVGWEVLAIGSPFGFDSVTAGIVSAKGRSL	197			
Qy	227	PNESYTFPIQTDVAINPONGSGPLFNLRKGVVGINSQIYSRGGFMGISFAIPIDVAMNY	286				
Db	198	PNESYVFPFIQTDVAINPONGSGPLNLEGEVVGINSQIFTRSGGFMGLSFAIPIDVAMNY	257				
Qy	287	AEQLKNTGKVGQQLGVIIQIEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIV	346				

[illegible]

```

RESULT      2
US-08-245-294-8
; Sequence 8, Application US/08245294
; Patent No. 5644047
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA
; TITLE OF INVENTION: INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/245,294
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-245-294-8

```

Query Match	31.1%	Score	731.5;	DB 1;	Length	503;			
Best Local Similarity	39.5%;	Pred. No.	5.2e-60;						
Matches	183;	Conservative	76;	Mismatches	173;	Indels	31;	Gaps	11;
QY	8	FALAAALACGACGAGSFFGADKKKASVFERIEHTKDGGSVSMLLPDRFAQLVQSEGPA	67						
Db	15	FSAALETALEFFSGC---GSSLWTTKAHANSV-----FSSLMOQOGFADIVSQVKPA	62						
QY	68	VNIOAAPAPRTQN---GSGNAETSDSDPLADSDP-----FYEEFKRLVPNNPEIQEEAADD	120						
Db	63	VSVSVQVKSNKKKEWFESDFSTPGFDQLDQHPKLFKFQDFDYNRDKPSNKS-L-QRSHRL	121						
QY	121	GGLNFGSGFTIISKNGLTNTWVAVGMSGTKVLLNDKREYTKALIGSDVSDVALLKIDA	180						
Db	122	RPAFGSGGFISSDGVIYTNHNVISDGTSAVVLDDGTTELNAKLIGDPTDRLAFLAVKNE	181						
QY	181	TEELPVVYKIGNPNKLPKGEVVAIGAAPPFGFDNSVTAGIVSAKGRSLPNSVTFPIQTDDVA	240						

Db	182	KRFESYVDFGDDSKLRVGDWVVAIGNPFLGGTVTACIYSARGRDI GTGVYDDFIQIDA	241
Qy	241	INPNSGGPLFNLIKQVGVINGINSOISYRSRGGFMGISFAIPDIDAMNVAEOLKNTGKVQRC	300
Db	242	VNRNMSGPTFDLNGKVGWGVNTAIFSPSGCNTGVIAEIPAATANEVVOOLIEFKGLVQRC	301
Qy	301	LGVIIQEVSYLQASQFGLDKASGALIAKTLPQSPAERAGLQAGDIVLSLDGGEIRSSGDL	360
Db	302	LGVOIQPVTKEISDSITGLKEAKCALITDPLKG-PAAKACIKAGDVIIISVNGEKINDVRDL	360
Qy	361	PVAVGAIPTCKEYSLGVWRKGEBITTAKALGNAAEHGTGASSKTDEAPYT-EOOSGTFSVE	419
Db	361	AKRIANNPGTETVLGVWKSCKREENIKVLDSPED- --ENMKDSKSYNEHCNSDETLE	417
Qy	420	SAGITITQHTDSSGKHLVVVV- --SDAAERAGLRHGDEILAV	459
Db	418	DYGLIIVAPSDDGUG- --LVVTDVDPDSDAAK- GIRPGDVIVTV	457

```

1  RESULT 3
2  US-08-474-499-8
3  ; Sequence 8, Application US/08474499
4  ; Patent No. 5693776
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Anderson, Burt E.
7  ; APPLICANT: Regnery, Russell L.
8  ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
9  ; TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
10 ; TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
11 ; NUMBER OF SEQUENCES: 10
12 ; NUMBER OF SEQUENCES: 10
13 ; CORRESPONDENCE ADDRESS:
14 ; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
15 ; STREET: 127 Peachtree Street, Suite 1200
16 ; CITY: Atlanta
17 ; STATE: Georgia
18 ; COUNTRY: USA
19 ; ZIP: 30303
20 ; COMPUTER READABLE FORM:
21 ; MEDIUM TYPE: Floppy disk
22 ; COMPUTER: IBM PC compatible
23 ; OPERATING SYSTEM: PC-DOS/MS-DOS
24 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
25 ; CURRENT APPLICATION DATA:
26 ; APPLICATION NUMBER: US/08/474,499
27 ; FILING DATE: 07-JUN-1995
28 ; CLASSIFICATION: 536
29 ; PRIOR APPLICATION DATA:
30 ; APPLICATION NUMBER: US 08/245,294
31 ; FILING DATE: 18-MAY-1994
32 ; ATTORNEY/AGENT INFORMATION:
33 ; NAME: Spratt, Gwendolyn D.
34 ; REGISTRATION NUMBER: 36,016
35 ; REFERENCE/DOCKET NUMBER: 1414.612
36 ; TELECOMMUNICATION INFORMATION:
37 ; TELEPHONE: 404/688-0770
38 ; TELEFAX: 404/688-9880
39 ; INFORMATION FOR SEQ ID NO: 8:
40 ; SEQUENCE CHARACTERISTICS:
41 ; LENGTH: 503 amino acids
42 ; TYPE: amino acid
43 ; TOPOLOGY: linear
44 ; MOLECULE TYPE: protein
45 ; US-08-474-499-8

```

```
Query Match      31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. NO. 5.2e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;
```

QY 68 VVNIQAAPAPRTQN---GSGNAETSDSDPLADSDP---FYEFFKRLVPMNPEIPOEADD 120  
Db 63 VVSQVKSNNKKKEWFFSDFSTPGFDQLPDQHPKLFQDFYNRKPKSNKSL-QRSHRL 121  
QY 121 GCLNFGSGFIISKNGYILTNTHVWAGMGSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180  
Db 122 RPIAFSGGFFISSDGYIVTNHNVISDGTSYAVVLDGTELNAKLIGTDPRTDLAVLKVNE 181  
QY 181 TEELPVWIKIGNPKLPGEWAAIAGAPFGDINSVTAGIVSAGRSPLPNSYTPFIQTDVA 240  
Db 182 KRKFYSYVDFGDDSKLRVGDWVAIGNPFLGGTVTVTAGIVSARGRDITGTGYDDFIQIDAA 241  
QY 241 INPGNSGGLFNLKGQVVGINSQIYSRSGFMGSIFAIPIDVAMNVAEOLKNTGKVQRGQ 300  
Db 242 VNRGNSGPTFDLNGKVVGVNTAIFSPSGNVGIAFAIPAATAEVVQQLIEKGLVQRGW 301  
QY 301 LGVLIQVSVYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
Db 302 LGVLIQVTKESISIGLKEAKGALITDPLKG-PAKAGIKAGDVIISVNGEKINDVRDL 360  
QY 361 PVMGAIPTGKESVGLVWRKGEIITIKAKLGNAAEHTGASSKTDEAPYT-EQOSGTFVSVE 419  
Db 361 AKRIANNSPGETVTLGVWKSKEENIKVKLDSMPED---ENMKDGSKYSNEHNSDETLE 417  
QY 420 SAGITLQTHDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 418 DYGLIVAPSDDGLG--LVVTDVDPDSDAADK-GIRPGDVIVTV 457  
RESULT 4  
US-08-307-279A-8  
; Sequence 8, Application US/08307279A  
; Patent No. 5736347  
; GENERAL INFORMATION:  
; APPLICANT: Anderson, Burt E.  
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae  
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea  
; TITLE OF INVENTION: Henselae and Rochalimaea Quintana Infection  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: 127 Peachtree Street, N.E., Suite 1200  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/307,279A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwendolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 1414.624  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 688-0770  
; TELEFAX: (404) 688-9880  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 503 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-307-279A-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;  
Best Local Similarity 39.5%; Pred. No. 5.2e-60;  
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;  
QY 8 FALAALCAALLAGCEKAGCFGADKKEASFVERIEHTKDDGVSVMILLPDPFQALVQSEGPA 67  
Db 15 FSALETALFTSGC---GSSLWTTKAHANSV-----FSSLMOQOQGFADIVSQVKPA 62  
QY 68 VVNIQAAPAPRTQN---GSGNAETSDSDPLADSDP---FYEFFKRLVPMNPEIPOEADD 120  
Db 63 VVSQVKSNNKKKEWFFSDFSTPGFDQLPDQHPKLFQDFYNRKPKSNKSL-QRSHRL 121  
QY 121 GCLNFGSGFIISKNGYILTNTHVWAGMGSIKVLLNDKREYTAKLIGSDVQSDVALLKIDA 180  
Db 122 RPIAFSGGFFISSDGYIVTNHNVISDGTSYAVVLDGTELNAKLIGTDPRTDLAVLKVNE 181  
QY 181 TEELPVWIKIGNPKLPGEWAAIAGAPFGDINSVTAGIVSAGRSPLPNSYTPFIQTDVA 240  
Db 182 KRKFYSYVDFGDDSKLRVGDWVAIGNPFLGGTVTVTAGIVSARGRDITGTGYDDFIQIDAA 241  
QY 241 INPGNSGGLFNLKGQVVGINSQIYSRSGFMGSIFAIPIDVAMNVAEOLKNTGKVQRGQ 300  
Db 242 VNRGNSGPTFDLNGKVVGVNTAIFSPSGNVGIAFAIPAATAEVVQQLIEKGLVQRGW 301  
QY 301 LGVLIQVSVYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360  
Db 302 LGVLIQVTKESISIGLKEAKGALITDPLKG-PAKAGIKAGDVIISVNGEKINDVRDL 360  
QY 361 PVMGAIPTGKESVGLVWRKGEIITIKAKLGNAAEHTGASSKTDEAPYT-EQOSGTFVSVE 419  
Db 361 AKRIANNSPGETVTLGVWKSKEENIKVKLDSMPED---ENMKDGSKYSNEHNSDETLE 417  
QY 420 SAGITLQTHDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 418 DYGLIVAPSDDGLG--LVVTDVDPDSDAADK-GIRPGDVIVTV 457  
RESULT 5  
PCT-US95-06211-8  
; Sequence 8, Application PC/TUS9506211  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: 127 Peachtree Street, Suite 1200  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06211  
; PRIOR APPLICATION NUMBER: US 08/245,294  
; FILING DATE: 18 MAY 1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwendolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 1414.6121  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404/688-0770  
; TELEFAX: 404/688-9880  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 503 amino acids

```

;      TYPE: amino acid
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
PCT-US95-06211-8

```

Query Match	31.1%;	Score 731.5;	DB 5;	Length 503;
Best Local Similarity	39.5%;	Pred. NO. 5.2e-60;		
Matches 183;	Conservative	76;	Mismatches 173;	Indels 31;
				Gaps 11;

QY	8	FALAAALCAALLACEKAGCFGGADKKEASFVERIEHTKDDGVSMILLPDPFAGLVQSEGPA	67
Db	15	FSAALETALFFSGC---GSSLWTTKAHANSV-----FSSLMOQOQFADIVSOVKPA	62
QY	68	VVNIOQAPAPRTON---GSGNAETSDSDPLADSDP---FVEEFKRLVPNNPEIPOEADD	120
Db	63	VVSVOVKSNNKKKEWFSEDFSTGPGFDQLPQDUHPLKFFQDFYNRDKPSNKSU-QRSHKL	121
QY	121	GGUNFGSGFTISKNGXYILTNTHVYVAGMSGIKVLLNDNKREYTKATLIGSDVQSDVALLKIDA	180
QY	122	RPIAFGSGGFFISSDGYIVTNHNVISDGTSVAVVDDPGTELNAKLIGTDPRTDLAVLKVNE	181
QY	181	TEELPVVKGIGNKPLKPEHVAIGAIPGFDNSVTAGIYSAGRSPLNPSYTFPTQTDVA	240
Db	182	KRRFSVDFDGDSSKLRRGVOMVAIGNPFGVLGGTVTAGIYSAGRDIGTGVYDDFIDQAA	241
QY	241	INPNSGGGLFNLKGOWGINSQIYSRSGGFMGISPAIPIDVAMNVAEOLKNTKVKVQRC	300
Db	242	VNRNGSGGPTDLNGKVVGVNTAIPSPSGNGVCIAPAIAPATANEVVQQLEIEKGLVQRC	301
QY	301	LGVIIQEVSYGLAQSGLDKASGALITAKILPQSPAPBRACLQAGDVIIVSLDGGIESSGDL	360
Db	302	LGVIQIPVTKESDSTGLKEARGALITDPLKG--PAAKAGIKAGDVIIVSYNGEKINDVRDL	360
QY	361	PVMVGAIITPGKEVSLGWRKRGEEITIKALGNAAEHTGASSKTDCAPIY--EQSGTFSVE	419
Db	361	AKRIANNSGCEVTVLGVWKSNGKEENIKVKLDSMPED---ENNKDGSKYSNEHGNSDTEI	417
QY	420	SAGITLOTHTDSSGKHLVVYRV---SDAAERAGLRHGDEILAV	459
Db	418	DYGLIIVAPSDDGIG--LVYTDVDPDSDAADK--GIRPGDVIVTV	457

```

RESULT      6
US-08-485-569-2
: Sequence 2, Application US/08485569
: Patent No. 5679547
:
: GENERAL INFORMATION:
: APPLICANT: Krivan, Howard C.
: APPLICANT: Samuel, James E.
: TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
: TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESS: Pennile & Edmonds
: STREET: 1155 Avenue of Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/485.569
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/903,079
: FILING DATE: 22-JUN-1992
: PRIOR APPLICATION DATA:

```

```

: APPLICATION NUMBER: US 07/810,966
: FILING DATE: 20-DEC-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/631,698
: FILING DATE: 21-DEC-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: Baldwin, Geraldine F.
: REGISTRATION NUMBER: 31,232
: REFERENCE/DOCKET NUMBER: 7969-049
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 463 amino acids
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-485-569-2

```

Query Match	27.5%;	Score 646;	DB 1;	Length 463;
Best Local Similarity	38.9%;	Pred. NO. 4.5e-52;		

Qy	54	LPDFAQLVQSEGPVAVNIQAAPRTQNGSGNAETDS-DPLADSDP--FYEFF-KRLVLPN	109
Db	28	LPSFVSQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDIIPEEFKFFCDRFAEQ	85
Qy	110	MPEIPEQEAADGGGLNFGSGFII--SKNGYILTNTHVAVMGCSIKVLLNDKREYTKAGLIGSD	168
Db	86	FGRGESKRNFRL--GSGVIIINASKGYVLTNHVIIDGADKITVQDQREFAKALGVGKD	143
Qy	169	VQSDVALLKIDAYEELPVWKIGNPKNLKPEEWAAICAPGPFNSVTAGIVSAKGRSLPN	228
Db	144	EQSDIALVQLEKPSNLTEIKFAUDSKLRVGDFVTAICNPGLCQTVTSGIVSALGRSTGS	203
Qy	229	ES--YTDFIOTDVAINPCNSGGPLNLKGVGVGINQIVSRSGGFMGISFAIPIDVAMNV	286
Db	204	DSGTENYIQTDAVNRNNGSGGALVNLGELIGINTAIIISPSCGNAGIAFAIPNQASNL	263
Qy	287	AEOLKNTKGVORGOLGVIIOEVSYGSLAOSPLDKASCALITAKTLPGSPERAGLQAGDIV	346
Db	264	VQQLTEFGQVRRGLTGKGLNLADAKAFNVSAQQGAFVSEVLPKSAEKKAGLKGADII	323
Qy	347	LSLDGGIRSSGDLPPVMVGAITPGKEVSLGVGRWKGBEITIKAKLGNAAEHTGASSKTEA	406
Db	324	TAMNGQKISSFAETRAKIATTGAGKEISLTRYLDGKSHDVKKML-QADDSQOLSSKT-EL	381
Qy	407	PYTEQSGSTSVESAGITLTQHTHDSGKHLVVVYV---SDAERAGLRHNGDEILAV	459
Db	382	PALD-----GATLKDYDAKGVKGEITFIKIQPNSLAAQR-GLKSGDIIFIIGI	425

RESULT 7  
US-08-480-993-2  
; Sequence 2, Application US/08480993  
; Patent No. 5721115  
; GENERAL INFORMATION:  
; APPLICANT: Krivan, Howard C.  
; APPLICANT: Samuel, James E.  
; APPLICANT: Samuel, Nils T.  
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE  
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.



```
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/903,079
; FILING DATE: 22-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,966
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,698
; FILING DATE: 21-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7969-050
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-480-993-2
```

```
Query Match 27.5%; Score 646; DB 1; Length 463;
Best Local Similarity 38.9%; Pred. No. 4.5e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

QY 54 LPDRAQLVQSGPAPVNIQAAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSPFLDDIPEEFKFFEGDRFAEQ 85
QY 110 MPEIQEADGGLNFGSGFTI-SKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSD 168
DB 86 FGGREGSKRNPRGL--GSGVIINASKGYVLTNNHVIDGADKITVQLQDGRFKAKLVGKD 143
QY 169 VQSDVALLKIDATELPVVKIGNPNKLPGEWVAIGAIPFGFDSNVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTETKFAADSKLRVGDFTVAIGNPFGQTVTSGIVSALGRSTGS 203
QY 229 ES--YTPFIQTDVAINPGNSGGLNKGQVVGINSQIYSRSGGFMGISFAIPIDVAMNV 286
DB 204 DSGYENYIQTDAAVNRGNSGALVNLGELIGINTAILSPSGNGAGIAFAPISNOASNL 263
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
DB 264 VQIILEFGQVRRLGILKGGELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAGLKAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDEA 406
DB 324 TAMNGOKISSAEIRAKIATTGAGREISLTYLRDCKSHDVKMKL-QADDGSQLSSKT-EL 381
QY 407 PYTEQSGTFSVESAGITLQTHDSSGHLVVVRV--SDAAERAGLRHGDITLAV 459
DB 382 PALD-----GATLKDYDAKGVKGIETITKQPNSLAAQR-GLKSGDITIGI 425
```

```
RESULT 8
US-09-079B-2
; Sequence 2, Application US/07903079B
```

```
; Patent No. 5843463
; GENERAL INFORMATION:
; APPLICANT: Krivan, Howard C.
; APPLICANT: Samuel, James E.
; APPLICANT: No 5843463berg, Nils T.
; TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
; TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/903,079B
; FILING DATE: 22-JUN-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,966
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/631,698
; FILING DATE: 21-DEC-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7969-004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-07-903-079B-2
```

```
Query Match 27.5%; Score 646; DB 2; Length 463;
Best Local Similarity 38.9%; Pred. No. 4.5e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

QY 54 LPDRAQLVQSGPAPVNIQAAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSPFLDDIPEEFKFFEGDRFAEQ 85
QY 110 MPEIQEADGGLNFGSGFTI-SKNGYILTNTHVVGMSIKVLLNDKREYTKALIGSD 168
DB 86 FGGREGSKRNPRGL--GSGVIINASKGYVLTNNHVIDGADKITVQLQDGRFKAKLVGKD 143
QY 169 VQSDVALLKIDATELPVVKIGNPNKLPGEWVAIGAIPFGFDSNVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTETKFAADSKLRVGDFTVAIGNPFGQTVTSGIVSALGRSTGS 203
QY 229 ES--YTPFIQTDVAINPGNSGGLNKGQVVGINSQIYSRSGGFMGISFAIPIDVAMNV 286
DB 204 DSGYENYIQTDAAVNRGNSGALVNLGELIGINTAILSPSGNGAGIAFAPISNOASNL 263
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346
DB 264 VQIILEFGQVRRLGILKGGELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAGLKAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGAITPGKEVSLGVWRKGEETIKAKLGNAAEHTGASSKTDEA 406
```

Db 324 TAMNQKISSPAERAKIATTGACKETSILYLRDGKSHDVNMKL-QADDSGSLSKT-EL 381  
QY 407 PYTQOQSGTTSVESAGITLOTHDSSGKHLVVVRV---SDAAERAGLRHGDDEILAV 459  
Db 382 PALD-----GATLKDYDAKGVKIEITKIOPNSLAAQR-GLKSGDIIIGI 425

## RESULT 9

US-08-350-741-2  
; Sequence 2, Application US/08350741  
; Patent No. 5804194  
; GENERAL INFORMATION:  
; APPLICANT: DOUGAN G.,  
; APPLICANT: CHARLES I.G.,  
; APPLICANT: HORMACHE C.E.,  
; APPLICANT: JOHNSON K.S.,  
; APPLICANT: CHATFIELD S.N.  
; TITLE OF INVENTION: LIVE VACCINES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON and VANDERHVE PC  
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/350,741  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 09-MAY-1994  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 30-NOV-1992  
; APPLICATION NUMBER: GB 9007194.5  
; FILING DATE: 30-MAR-1990  
; APPLICATION NUMBER: PCT/GB91/00484  
; FILING DATE: 28-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILSON, MARY J.  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: amino acids  
; LENGTH: 475 amino acids  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-350-741-2

Query Match 26.9%; Score 633.5; DB 1; Length 475;  
Best Local Similarity 34.1%; Pred. No. 7e-51;  
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

QY 44 TKDGSVSMLLPFAQLVQSEGPAVV--NIQAAPAPRTONGSGNAE----TDSPLADSD 97  
Db 29 TSSSMTAQQNPSPALPMEKVMKSVINVEGSTVTVPMPRNPFQFFGDDSPFCQDGS 88  
QY 98 PFYEYFKRLVPNPEIPQEEADGG-----LNFGSGFII-SKNGYILNTNTHVVMGM 148  
Db 89 PF-----QNGSPFC--QGGGNGGQOQKFMALGSGVVIDAAKGVVVTNNHVVDNAS 138

QY 149 SIKVLNDKREYTAKLIGSDVQSDVALLKIDATEELPWWKIGNPKNLKPGEWAAIGAPF 208  
Db 139 VIKVOLSDGRKFDKAVGVGKDPDSIALIQIONPKNLTAIKLADSDALRVGDTTVAIGNPF 198  
QY 209 GFDNSVTAGIVSAKGRS-LPNESYTPPIQTDVAINPNSGGPLPNLKGQVVGINSQIYSR 267  
Db 199 GLGETVTSGIVSALGRSGLNVENFIQTDAAINRSGALVNLNGELIGINTAILAP 258  
QY 268 SGGFMGISFAIPIDVAMNVAEQLKNTCKVORGOLGVIIQEVSYGLAQSFGLDKASGALIA 327  
Db 259 DGNIGIGFAIPSNMVKNLTSQMVEYGVQVRKGLGIMGTLSNELAKAMKVDAGRAPVS 318  
QY 328 KILPGSPAERAGLQAGDIVLSLDGGEIRSSGDLPMVMVGAITTPGKEVSLGVNMRKGEETIK 387  
Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFAALRAQVGTWPVGSKISLGLLREGKAITVN 378  
QY 388 AKLGNAAHTGASSKTDAPYTEQOQSGTFSVESA-----GITPLOTHTDSSGKHLVVVR 440  
Db 379 LEL-----QSSQSOQVDSSTIFSGIEGAEKMSKNGODKGVVWSSVK 418  
QY 441 VSDAAERAGLRHGDDEILAVRASP 463  
Db 419 ANSPAAQIGLKKGDVIGANQQP 441

## RESULT 10

US-08-463-875A-2  
; Sequence 2, Application US/08463875A  
; Patent No. 5980907  
; GENERAL INFORMATION:  
; APPLICANT: DOUGAN, Gordon  
; APPLICANT: CHARLES, Ian G.  
; APPLICANT: HORMACHE, Carlos E.  
; APPLICANT: JOHNSON, Kevin S.  
; APPLICANT: CHATFIELD, Steven N.  
; TITLE OF INVENTION: LIVE VACCINES  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON and VANDERHVE PC  
; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD  
; CITY: ARLINGTON  
; STATE: VIRGINIA  
; COUNTRY: USA  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/463,875A  
; FILING DATE: 05-JUN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/340,741  
; FILING DATE: 07-DEC-1994  
; APPLICATION NUMBER: US 07/952,737  
; FILING DATE: 30-NOV-1992  
; APPLICATION NUMBER: GB 9007194.5  
; FILING DATE: 30-MAR-1990  
; APPLICATION NUMBER: PCT/GB91/00484  
; FILING DATE: 28-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WILSON, MARY J.  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-158  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 475 amino acids

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-463-875A-2

```

Query Match 26.9%; Score 633.5; DB 2; Length 475;  
Best Local Similarity 34.1%; Pred. No. 7e-51;  
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

RESULT 11  
US-08-278-091-2  
Sequence 2, Application US/08278091  
Patent No. 5506139  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: KOEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with  
NUMBER OF INVENTION: Reduced Protease Activity  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/278,091  
FILING DATE: 21-JUL-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:

```

: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24 973
: REFERENCE/DOCKET NUMBER: 1038-371
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 463 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-278-091-2

```

Query Match	26.9%;	Score 632;	DB 1;	Length 463;
Best Local Similarity	38.5%;	Pred. No. 9.3e-51;		
Matches 160;	Conservative 75;	Mismatches 153;	Indels 28;	Gaps 12;
QY	54	LPDFAQLVQSEGPAVVNIQIAPAPRTONGSNAETDS-DPLADSDP--FYEFF-KRLVPN	109	
Db	28	LPFEVSEQNSLAPMLEKVQ--PAVVTLVSFGKAKVDSPFLDDIPPEEKKFFFGDRFAEQ	85	
QY	110	MPEIPOEEADGGFLNFGSGRII-SKNGYILTNTHVAVGMSIKVLLNDKREYTKALLTCS	168	
Db	86	FGRGESKRNFRL--GSGVIIWASKGYVLTNHEVTDADKIYVQLQDGRFKAIVGKD	143	
QY	169	VQSDVALLKTDATFELPVYKIGNPKNLKPGEWAAAGAPFGDNSVTAGISVAKGRSLPN	228	
Db	144	EISDIALVQLKPSNTEIKFADSDRLVRGDFTVAIIGNPFLGTQTVISVSLALRGSTG	203	
QY	229	ES--YTPFIQTDVAINPGNSGGFLNFLKGQVGINSQIYVSRSGFGWGISFAIPDIVAMNV	286	
Db	204	DSGTENYIQTDAAVNRKNSGGALVNLGELIGINTAIISPSCGNAGIAIFAISNQASNL	263	
QY	287	ASOLKNTGKVRQGLGVITQEVSYGLAQSFGLDKASGALIAKTLPGSPAERAGLQAGDIY	346	
Db	264	VQOILFEQVRRGLGIKGELNADIAKAFNVSQAQGFVSEVLPRPSAAEKAGLKAGDII	323	
QY	347	LSLDGGETIRSSGDLPMVWGALTPGKEVSLGVWRKGBEITIKAKLGNAAEHTGASSKTEA	406	
Db	324	TAMNGQKISSFAEIRAKATTGAGREISLTLRDCKSHDVKKML-QADSSQLSKYT-EL	381	
QY	407	PYTEQSGGTFSVESAGITLQTHDTSGRKHLVVVR---SDAAERAGLRHGDIELAV	459	
Db	382	PALD-----GATLKDTAKGVKGIEITKTQPNSLAOR-GLKSGDIIIGI	425	

```

RESULT 12
US-08-483-859-2
; Sequence 2, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/483,859
: FILING DATE: 07-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/296,149
: FILING DATE: 26-AUG-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/278,091
: FILING DATE: 21-JUL-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I.
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 463 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-483-859-2

```

Query Match	26.9%	Score 632:	DB 1:	Length 463:
Best Local Similarity	38.5%	Pred. No. 9.3e-51:		
Matches 160:	Conservative 75:	Mismatches 153:	Indels 28:	Gaps 12:
QY	54	LPDFAQLVQSEGPAVVNIQAAPRTONGSGNAETDS-DPLADSDP--FYEFF-KRLVNP	109	
Db	28	LPSEVSEQNSLAPMLEKVQ--PAVVTLSVGKAKVDSRSPFLDDIPREFFFGDRFAEQ	85	
QY	110	MPEIQBEADGGGLGSGFGPII-SKNGYIILTTHVAVAGMSIKYLLNDKREYTKAGL	168	
Db	86	FGRGESKRNFRL--GSGVIINASKGYVLTNNHVIDEADKITVOLQDGRFEKAKLVGKD	143	
QY	169	VQSDVALLKTDATTELPPVKIGNPKNLKPGEWAACAPGCFDMSVYTAGIVSAKGRSLPN	228	
Db	144	ELSDIALVQLEKPSNLTEIKFADSKURVGDTVAIGNPFLGQGTVTSGVIVSALGRSTGS	203	
QY	229	ES--YTFPIQTDVAINPNSGGPLFLNLKGQVVGINSQIYRSRGGFMGISFAIPIDVAMNV	286	
Db	204	DSGTENYIQTDAAVNRNGSGGALVNLGELIGINTAIISPSGGNAGIATPAIPSNQASNL	263	
QY	287	AEQLKNTGKVQRCGLGVIIOEVSYGLAQSGFLDKASGALATAKILPGSPAERAGLQAGDIV	346	
Db	264	VQOILEFGQVRRGLGKGGELNADLAKAFNVSAQOAGAFVSEVLPSKAAEKAGLKAGDII	323	
QY	347	LSLDGGIRSGDLPVVMVGAITPGCKEVSGLGWKRGEITIKAKLGNAEHTGASKKIDEA	406	
Db	324	TAMNQKISSFAETRAKIATGTAGKELISLYLRDCKSHDVNMKL-QADDSQLSSKT-EL	381	
QY	407	PYTEQSGTFSVESAGITLTQHTDSSGKHLVVVRV---SDAAERAGLRHGDILAV	459	
Db	382	PALD-----GATLKDYDAKGVKYEITIKIOPNSLAAOR-GUKSGDIIIGI	425	

```

RESULT 13
US-08-472-173-2
; Sequence 2, Application US/08472173
; Patent No. 565533
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M
; APPLICANT: YANG, Yan-ping
; APPLICANT: CHONG, Pele
; APPLICANT: OMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of H
; TITLE OF INVENTION: Reduced Pro
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:

```

```

ADDRESS: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,173
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-493 MIS:v9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO.: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-472-173-2

```

Query Match 26.9%; Score 632; DB 1; Length 463;  
Best Local Similarity 38.5%; Pred. NO. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels

[illegible]

RESULT 14  
US-08-487-167-2  
; Sequence 2, Application US/08487167

Patent No. 5869302  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M.  
APPLICANT: YANG, Yan-Ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hin7 Protein with  
REDUCED PROTEASE ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487.167  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296.149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278.091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-167-2

Query Match 26.9%; Score 632; DB 2; Length 463;  
Best Local Similarity 38.5%; Pred. No. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;  
QY 54 LPDFAQLVQSGPAAVNIQAAPRTQNGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVTLVSECKAKVDSRSPFLDDIPPEPKFFGDFAEQ 85  
QY 110 MPEIQEADGGNFGSGFII-SKNGYILFNTHVAGMGSIKVLLNDKREYAKLIGSD 168  
DB 86 FGGRGESKRNRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFKAKLVGKD 143  
QY 169 VQSDVALLKIDATELPPVVKIGNPNKLPGEVVAIGAPFGFDSNVTAGIVSAKGRSLPN 228  
DB 144 ELSDIALVQLEKPSNLTETKFAFADSKLRVGDFTVAIGNPFGLGQTVTSIGVSAKGRSGS 203  
QY 229 ES--YTPFIQTDVAIPNCGSGPLNKGQVVGINSOISYRSGGFMGIFAPIDVAMNV 286  
DB 204 DSGTYENTQDAVNRNCGALVNLGELIGINTAIISPGCGNAGIAFAIPNSQASNL 263  
QY 287 AEQLANTKVRQGLQVITQIYESYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
DB 264 VQOILEFGVRRGLIGKGGELNADLAKAFNVSAQOGAFVSEVLPKSAEAKGLKAGDII 323  
QY 347 LSLDGGEIRNSGDLPMVVGATTPGKEVSLGVVRKGEETIKAKLGNAAEHTGASSKTEA 406

Db 324 TAMNGQISSFAEIRAKIATTGAGKEISLTLYLRDCKSHDVKNKL-QADSSQLSKT-EL 381  
QY 407 PYTEQSQSTVESAGITLQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 382 PALD-----GATLKDYDAKGVKGIEITKTOPNSLAAQR-GLKSGDIIIGI 425  
RESULT 15  
US-08-482-816-2  
Sequence 2, Application US/08482816  
Patent No. 5935573  
GENERAL INFORMATION:  
APPLICANT: LOOSMORE, Sheena M  
APPLICANT: YANG, Yan-ping  
APPLICANT: CHONG, Pele  
APPLICANT: OOMEN, Raymond P.  
APPLICANT: KLEIN, Michel H.  
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease  
ACTIVITY  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/482.816  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/296.149  
FILING DATE: 26-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/278.091  
FILING DATE: 21-JUL-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I.  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 463 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-482-816-2

Query Match 26.9%; Score 632; DB 2; Length 463;  
Best Local Similarity 38.5%; Pred. No. 9.3e-51;  
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;  
QY 54 LPDFAQLVQSGPAAVNIQAAPRTQNGSGNAETDS-DPLADSDP--FYEFF-KRLVNP 109  
DB 28 LPSFVSEQNSLAPMLEKVQ--PAVTLVSECKAKVDSRSPFLDDIPPEPKFFGDFAEQ 85  
QY 110 MPEIQEADGGNFGSGFII-SKNGYILFNTHVAGMGSIKVLLNDKREYAKLIGSD 168  
DB 86 FGGRGESKRNRGL--GSGVIINASKGYVLTNNHVIDEADKITVQLQDGRFKAKLVGKD 143  
QY 169 VQSDVALLKIDATELPPVVKIGNPNKLPGEVVAIGAPFGFDSNVTAGIVSAKGRSLPN 228  
DB 144 ELSDIALVQLEKPSNLTETKFAFADSKLRVGDFTVAIGNPFGLGQTVTSIGVSAKGRSGS 203

QY 229 ES--YTPFIQTDVAINPGNSGGPLFNLKGQVVGINSQIYSRSGGFMGISFAIPIDVAMNV 286  
:| | :|||| | :|||| | | :||| : | | | :||| : | | :  
Db 204 DSGTYENYIQTDAAVNRGNSGGALVNLNGELIGINTAIISPGGNAGIAFAIPSNQASNL 263  
QY 287 AEOLKNTCKVORGOLGVIIQEVSYCLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIV 346  
:| : | :||| | : | :||| : | :||| : | :||| : | :||| : | :||| :  
Db 264 VQQILEFGQVRRGLGIGKGGELNADLAKAFNVSAQQGAFVSEVLPKSAAEKAGLKAGDII 323  
QY 347 LSLDGEIRSSGDLPLVMVGALTGPKEVSLGVWRKGEITIRAKLGNAAEHTGASSKTDEA 406  
:| : | :||| : | :||| : | :||| : | :||| : | :||| : | :||| :  
Db 324 TAMNGQKISSFAEIRAKIATTGACKEISLTLRDGKSHDVNKL-QADDSSOLSSKT-EL 381  
QY 407 PYTEQQSGTFSVESAGITLQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
| : | :||| : | :||| : | :||| : | :||| : | :||| : | :||| :  
Db 382 PALD-----GATLKDYDAKGVKGIEITKIQPNSLAAQR-GLKSGDIIIGI 425

Search completed: June 10, 2002, 12:10:53  
Job time: 92 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:21 ; Search time 20.95 Seconds  
(without alignments)  
2132.770 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALAALCAALLAG.....ERAGLRHGDILAVRASPRQ 465  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	2 B81914	probable periplasm
2	985.5	41.9	474	2 F83550	serine proteinase
3	862.5	36.7	514	2 A82581	periplasmic protei
4	735	31.2	523	2 A97479	probable serine pr
5	735	31.2	523	2 A12696	serine proteinase
6	716	30.4	493	2 C97605	probable serine pr
7	716	30.4	514	2 A82827	serine proteinase
8	693.5	29.5	513	2 AD3418	serine proteinase
9	692.5	29.4	513	2 I40060	serine proteinase
10	683	29.0	497	2 B81728	serine proteinase,
11	683	29.0	497	2 H71465	probable do serine
12	683	29.0	524	2 AG3328	proteinase do (EC
13	669.5	28.5	488	2 G81528	serine proteinase,
14	668.5	28.4	488	2 H86612	DO serine proteina
15	668.5	28.4	488	2 G72011	do serine proteina
16	653.5	27.8	457	2 AG0433	proteinase (EC 3.4
17	649	27.6	481	2 AH0410	global stress requ
18	648	27.5	466	1 A64113	heat shock protein
19	639	27.2	474	2 A13349	proteinase DO (EC
20	638	27.1	511	2 F97720	periplasmic serine
21	637	27.1	474	1 I40059	htrA-like protein
22	636.5	27.1	475	2 AC0528	protease DO precu
23	636	27.0	530	2 F87590	serine proteinase
24	633.5	26.9	475	1 S15337	heat shock protein
25	632.5	26.9	456	2 F82307	proteinase DO VC05
26	631	26.8	429	2 AD1894	serine proteinase
27	627.5	26.7	455	2 JC6051	trypsin-like prote
28	623.5	26.5	455	2 C91142	serine endoprotein
29	617.5	26.2	455	2 F85987	serine endoprotein

30	617.5	26.2	455	2 AB0909	serine protease (E
31	615.5	26.2	513	2 B71722	probable periplasm
32	614.5	26.1	459	2 F72359	periplasmic serine
33	610.5	25.9	474	2 S45229	proteinase DO (EC
34	610.5	25.9	474	2 E85500	proteinase DO (EC
35	610.5	25.9	474	2 E90649	proteinase DO (EC
36	606.5	25.8	452	2 S77538	serine proteinase
37	604.5	25.7	481	2 D82826	heat shock protein
38	600.5	25.5	416	2 S75445	proteinase hhoB (E
39	594.5	25.3	478	2 G84956	proteinase do prec
40	589	25.0	389	2 B83089	AlpW protein PA444
41	582	24.7	476	2 H71936	proteinase DO - He
42	574.5	24.4	443	1 C64647	serine proteinase
43	571	24.3	453	1 B70426	periplasmic serine
44	564	24.0	468	2 AI2811	serine proteinase
45	564	24.0	495	2 B97590	htrA protein homol

ALIGNMENTS

RESULT 1

B81914

probable periplasmic serine proteinase (EC 3.4.21.-) NMA0710 [imported] - Neisseria

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: B81914

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.;

R:Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand

Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z24

A:Reference number: A81775; MUID:20222556

A:Accession: B81914

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-499 <PAR>

A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB83996.1; PID:g7

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA0710

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

Query Match	95.5%;	Score	2246;	DB 2;	Length	499;			
Best Local Similarity	95.9%;	Pred. No.	4.6e-138;						
Matches	444;	Conservative	6;	Mismatches	13;	Indels	0;	Gaps	0;

Qy	1	VFKKYQYFALAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMLLPDFVQL	60
Db	1	MFKKYQYLALCAALCAASLAGCDKAGSFFGADKKEASFVERIKHTKDDGSVSMLLPDFVQL	60
Qy	61	VQSEGPAVNTQAPAPRTQNGSNAETSDPLADSDPFYEFFKRLVPMNPPEIQQEADD	120
Db	61	VQSEGPAVNTQAPAPRTQNGSNAETSDPLADSDPFYEFFKRLVPMNPPEIQQEADD	120
Qy	121	GGLNFGSGFIISKNGYILTNTHVWAGMGSIKVLLNDKREYTKAKLIGSDVQSDVALLKIDA	180
Db	121	GGLNFGSGFIISKDGYILTNTHVVTGMGSIKVLLNDKREYTKAKLIGSDVQSDVALLKIDA	180
Qy	181	TEELPVVKIGNPKNLKPGEWAAIAGPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDDVA	240
Db	181	TEELPVVKIGNPKDLKPGEWAAIAGPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDDVA	240
Qy	241	INPNSGGPLFNKLGQVVGINSQIYSRSGGFMGISFAIPIDVAMNVAEQLKNTGVQRGQ	300
Db	241	INPNSGGPLFNKLGQVVGINSQIYSRSGGFMGISFAIPIDVAMNVAEQLKNTGVQRGQ	300
Qy	301	LGVIIEQVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL	360
Db	301	LGVIIEQVSYGLAQSGFLDKAGGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL	360
Qy	361	PVMVGAITPGKEVSLGVWRKGEIITKAKLGNAAEHTGASSKTDAPYTEOOSCTFSVES	420

Db 361 PVMVGAIPTPKESVSLGVWRKGEETIKVLCGNAAEHICASSKTDPAPTYEQQSFTSVES 420  
|||||

Qy 421 AGITLQHTDSSGKHLVVRVSDAAERAGLRHGDDEILAVRAS 463  
|||||

Db 421 AGITLQHTDSSGKHLVVRVSDAAERAGLRHGDDEILAVGVP 463  
|||||

#### RESULT 2

F83550  
serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: F83550

R:Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

.; Lozy, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho

Reference number: A89950; MUID:20437337

C:Accession: F83550

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-474 <STO>

A:Cross-references: GB:AE004511; GB:AE004091; NID:99946646; PIDN:AAG04155.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: mucD; PA0766

Query Match 41.9%; Score 985.5; DB 2; Length 474;

Best Local Similarity 52.5%; Pred. No. 2.1e-56;

Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

Qy 54 LPDFAOLVQSEGPVAVNI---QAAPRTQNGSGNAETSDPLADSP-FYEFKRLVFN 109  
|||||

Db 27 LPDFTPLVEQASPAVNISTRQKLP----DRAMRGQLSLDLEGLPMPFRDFLERS:IPQ 82  
|||||

Qy 110 MPEIP---QEEADGGLNGSGFIISKNGYILTNTHVVGAGMSIKVLLNDKREYTA 166  
|||||

Db 83 VPRNPRGQREQA---SLGSGFIISNDGYILTNHHVVADADETLVRLSDRSEHAKLIG 138  
|||||

Qy 167 SDVQSDVALLKIDATELPPVVKIGNPKNKPGEWVAAGPFGDINSVTAGIVSAKRSL 226  
|||||

Db 139 ADPRSDVALLKIDK-KNLPFLTKLGDNSNKLKVGWVAIGSPFGDHSVTAGIVSAKRSL 197  
|||||

Qy 227 PNEYTPFTQTDVAIPNGNSGGLFLNLKGQVVGINSQIYRSRSGFMGISFAIPIDVAMNV 286  
|||||

Db 198 PNEYTPFTQTDVAIPNGNSGGLFLNLQGEVVGINSQIFTRSGGFMGLSFAIPIDVAMNV 257  
|||||

Qy 287 ADQLKNTGKVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346  
|||||

Db 258 ADQLKNTGKVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 317  
|||||

Qy 347 LSLDGGGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEETIKVLCGNAAEHICASSK 405  
|||||

Db 318 LSLNGOSINESADPLHLVGNMKPKDINLDVIRNGQRKSLSMVAGSLPD-----DDEE 370  
|||||

Qy 406 ----APYTEQQSGTFSVESAGITLQTH--TDSGKHLVVRVSDA-AERAGLRHGDDEI 456  
|||||

Db 371 IASMGAPGAERSSNRLGVTVVADLTAEQRKSLDIQGGVVIKEVQDGPAAVIGLRPGDVI 428  
|||||

#### RESULT 3

A92581

periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Sep-2000

C:Accession: A82581

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A92581

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <SIM>

A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAF85040.1; GSPDB:G

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrato, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; F

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; L

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Marti

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C

.F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Saw

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Sil

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2241

C:Superfamily: Helicobacter serine proteinase

Query Match 36.7%; Score 862.5; DB 2; Length 514;

Best Local Similarity 46.2%; Pred. No. 2.2e-48;

Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;

Qy 54 LPDFAOLVQSEGPVAVNIOAAPRTQNGSGNAETSDPLADSDPPYEFKRL-----V 107  
|||||

Db 54 LPDFTQLVDQVGPVVNIETVITRKVKGRRIPLDND-----IPEFPRRFFGPDFQM 106  
|||||

Qy 108 PNWPEIPEEADGGGL---NFGSGFIISKNGYILTNTHVVGAGMSIKVLLNDKREYTA 164  
|||||

Db 107 PNQPRGGQD---DEGGTAGRGMSGGFIISKDGYILTNHHVITGTAASEVTIKUTDRREKAKI 164  
|||||

Qy 165 IGSVDQSDVALLKIDATELPPVVKIGNPKNKPGEWVAAGPFGDINSVTAGIVSAKR 224  
|||||

Db 165 IGSDEQYDVALLKIDA-KNLPVTRIGDSSLSKSGQWVAIGSPFGLDHSVTAGIVSALGR 223  
|||||

Qy 225 SLPNRS-YTPFTQTDVAIPNGNSGGLFLNLKGQVVGINSQIYRSRSGFMGISFAIPIDVA 283  
|||||

Db 224 STSDDRYVPFTQTDVPIQNGNSGGLFLNLTRGEVIGINSQIFSASGSGMISFAIPINLA 283  
|||||

Qy 284 MNVAEQKNTGKVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343  
|||||

Db 284 INAAEQIRKTKGVRQGLGVITQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343  
|||||

Qy 344 DIVLSLDGGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGE--ETIKAKLGNAAEHTGASS 401  
|||||

Db 344 DVIRSVNGKVISSFDLPPLICMPPGPKATLGIIRDGKPREIVTVTLTSLALN-QDATASED 402  
|||||

Qy 402 KTEAPYTEQQSGTFSVESAGITLQTHDSSSKHLVVRVSDAAERAGLRHGDDEILAVRA 461  
|||||

Db 403 EDDTAPSKPETS--NVELLGLQVNLNSAERERL-----ASSQNAKGGVRRITAVTA 452  
|||||

#### RESULT 4

A97479

probable serine proteinase homolog precursor [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002

C:Accession: A97479

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Gol

A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacteriu

A:Reference number: A97359; PMID:11743194

A:Accession: A97479

A:Status: preliminary

A:Molecule type: DNA



A;Residues: 1-523 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK86786.1; PID:g15155988; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C.1792  
A;Map position: circular chromosome

Query Match 31.2%; Score 735; DB 2; Length 523;  
Best Local Similarity 37.0%; Pred. No. 4.1e-40;  
Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

```
QY 36 SFVERIEHTKDDGVSMLLPDFAQLVQSEGPVAVNIOAAPRTQNGSGNAETDS----- 90
Db 44 SFAAPVEVTAPO-----VPSFANVVDVAVPAVVS-----RVQSNVQPADSSNSFSF 91
QY 91 -----DPLADSDPFYEFFKRLVPMNPEIQEADGGGLN-----FGSGFI 130
Db 92 NFGGRGLDQLPDDHPLKRFKKEFGQ-----NODRSRGPNRHRDGGKGLRPVAGSGGFF 146
QY 131 ISKNGYILTNTHVAVMGSTKVLNDRKREYAKLIGSDVQSDVALLKIDATEELPVVKIG 190
Db 147 ISEGGYVVTNNHVVDDGSAYTVVMNDGTELEAKLVGRDPRDLDALLKVDVNRKFTYVKFA 206
QY 191 NPNLKPGEVAAIGAIPFGFDSNVTAGIVSAKGRSLPNESTPFIQTDVAINPNSGGPL 250
Db 207 DDTKIRGVDMVAVGNPFGGLGTVTSGIISARGDRIGSGPYDDYQIDAAVNRNMSGGPA 266
QY 251 FNLKGVVGINSOIYRSRSGFMGISFAIPIDVAMNVAEQKNTKGQVORGOLGVITQEVSY 310
Db 267 FNLNGEVVGINTAIFSPSGGNVGIAPSPSVAKVDIADLQKDGKVERGWLGVQIQPVSK 326
QY 311 GLAQSGFLDKASGALLAKILPGSPAERAGIQAQDIVLSLDGGETRSSGDLPMVVGATPG 370
Db 327 DIAESLGLAEAKGALVVSQSGSPGKAGIKQGDIIITAVNGDPVKDARDLSRRIGGMAPN 386
QY 371 KEVSLGWVRKGEITIKAKLGN--AAEHTGASSKTDEAPYTEQSGTFSVESAGITLQTH 428
Db 387 SKVEISLWRGKGSQSVTVTLGDLTSDASKATPSQNDKKGSSSEKVLSSGLTVSPS 446
QY 429 TDSGKHLVVVRV---SDAAERAGLRHGDILAV 459
Db 447 DDGNG--LAITDVPDSDAAAR--GLTKGKITSV 477
```

RESULT 5

serine proteinase DO-like proteinase dop [imported] - Agrobacterium tumefaciens (strain  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C;Accession: A12696  
A;Authors: Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;  
Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,  
Science 294, 2317-2323, 2001  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; PMID:11743193  
A;Accession: A12696  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-523 <KUR>  
A;Cross-references: GB:AE008688; PIDN:AAL41991.1; PID:g17739363; GSPDB:GN00186  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: dop  
A;Map position: circular chromosome

Query Match 31.2%; Score 735; DB 2; Length 523;  
Best Local Similarity 37.0%; Pred. No. 4.1e-40;  
Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

```
QY 36 SFVERIEHTKDDGVSMLLPDFAQLVQSEGPVAVNIOAAPRTQNGSGNAETDS----- 90
Db 44 SFAAPVEVTAPO-----VPSFANVVDVAVPAVVS-----RVQSNVQPADSSNSFSF 91
QY 91 -----DPLADSDPFYEFFKRLVPMNPEIQEADGGGLN-----FGSGFI 130
Db 92 NFGGRGLDQLPDDHPLKRFKKEFGQ-----NODRSRGPNRHRDGGKGLRPVAGSGGFF 146
QY 131 ISKNGYILTNTHVAVMGSTKVLNDRKREYAKLIGSDVQSDVALLKIDATEELPVVKIG 190
Db 147 ISEGGYVVTNNHVVDDGSAYTVVMNDGTELEAKLVGRDPRDLDALLKVDVNRKFTYVKFA 206
QY 191 NPNLKPGEVAAIGAIPFGFDSNVTAGIVSAKGRSLPNESTPFIQTDVAINPNSGGPL 250
Db 207 DDTKIRGVDMVAVGNPFGGLGTVTSGIISARGDRIGSGPYDDYQIDAAVNRNMSGGPA 266
QY 251 FNLKGVVGINSOIYRSRSGFMGISFAIPIDVAMNVAEQKNTKGQVORGOLGVITQEVSY 310
Db 267 FNLNGEVVGINTAIFSPSGGNVGIAPSPSVAKVDIADLQKDGKVERGWLGVQIQPVSK 326
QY 311 GLAQSGFLDKASGALLAKILPGSPAERAGIQAQDIVLSLDGGETRSSGDLPMVVGATPG 370
Db 327 DIAESLGLAEAKGALVVSQSGSPGKAGIKQGDIIITAVNGDPVKDARDLSRRIGGMAPN 386
QY 371 KEVSLGWVRKGEITIKAKLGN--AAEHTGASSKTDEAPYTEQSGTFSVESAGITLQTH 428
Db 387 SKVEISLWRGKGSQSVTVTLGDLTSDASKATPSQNDKKGSSSEKVLSSGLTVSPS 446
QY 429 TDSGKHLVVVRV---SDAAERAGLRHGDILAV 459
Db 447 DDGNG--LAITDVPDSDAAAR--GLTKGKITSV 477
```

RESULT 6

probable serine proteinase DO-like precursor [imported] - Agrobacterium tumefaciens  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C;Accession: C97605  
A;Authors: A.; Liu, F.; Winkler, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Gol  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium  
A;Reference number: A97359; PMID:11743194  
A;Accession: C97605  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-493 <KUR>  
A;Cross-references: GB:AE007869; PIDN:AAK87796.1; PID:g15157169; GSPDB:GN00169  
C;Genetics:  
A;Gene: AGR\_C.3700  
A;Map position: circular chromosome

Query Match 30.4%; Score 716; DB 2; Length 493;  
Best Local Similarity 36.4%; Pred. No. 6.5e-39;  
Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

```
QY 48 GSVSMLLPDFAQLVQSEGP-----AVNIOAAPRTQNGSGNAETDSPLAD 95
Db 3 GSLS---APVAARAGSHGPESVADLAEPDLDAVNLISTQNVKTE--GKGPV---PKLPE 55
QY 96 SDPYEFPKRLVPMNPEIQEADGG---LNFSGGFIISKNGYILTNTHVAVMGSIKV 152
Db 56 GSPFQEFK-----DYFDSOKPEGKVNLSGLSGFIDPAGYVVVNNHVIAGDAIEV 108
QY 153 LINDKREYAKLIGSDVQSDVALLKIDATEELPVVKIGNPKNLKPGEWAAAGPFGDN 212
Db 109 IFPNGSKLAKTLVGTDTKTLSLVKVEPKTPLKAVKFGDSRMRIGDWMVAVNPGFLGG 168
QY 213 SVTAGIVSAKGRSLPNESTPFIQTDVAINPNSGGPLFNLKGVVGINSOIYRSRSGGM 272
Db 169 SLTVGVTVSARGNINAGPYDNFIQTDALINKNSGGPLFNMKGVEVIGINTAISPSSGSI 228
```



[illegible]



RESULT 15  
G72011  
do serine proteinase - Chlamydophila pneumoniae (strain CWL029)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: G72011  
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72011  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <ARN>  
A:Cross-references: GB:AE001678; GB:AE001363; NID:g4377301; PIDN:AAD19116.1.  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase

Search completed: June 10, 2002, 12:09:52  
Job time: 31 sec

RESULT 15  
G72011  
do serine proteinase - Chlamydophila pneumoniae (strain CWL029)  
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: G72011  
R:Kalmán, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72011  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <ARN>  
A:Cross-references: GB:AE001678; GB:AE001363; NID:g4377301; PIDN:AAD19116.1; PID:g437730  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:10:31 ; Search time 13.48 seconds  
(without alignments)  
1335.652 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALAALCAALLAG.....ERAGLRHGDEILAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 segs, 38719550 residues  
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	747	31.7	504	1	DEGP_RHIME
2	731.5	31.1	503	1	DEGP_BARHE
3	692.5	29.4	513	1	DEGP_BRUAB
4	683	29.0	497	1	DEGP_CHLMU
5	683	29.0	497	1	DEGP_CHLTR
6	668.5	28.4	488	1	DEGP_CHLPN
7	648	27.5	466	1	HTOA_HAEIN
8	633.5	26.9	475	1	DEGP_SALTY
9	627.5	26.7	455	1	DEGP_ECOLI
10	615.5	26.2	513	1	DEGP_RICPR
11	610.5	25.9	474	1	DEGP_ECOLI
12	594.5	25.3	478	1	DEGP_BUCAL
13	571.5	24.3	478	1	DEGP_BUCAP
14	470	20.0	437	1	DEGP_ARATH
15	448	19.0	480	1	HRA1_HUMAN
16	447.5	19.0	355	1	DEGS_ECOLI
17	436.5	18.6	413	1	HTRA_LACHE
18	435.5	18.5	480	1	HRA1_MOUSE
19	435	18.5	448	1	DEGS_ARATH
20	428	18.2	340	1	DEGS_HAEIN
21	414.5	17.6	458	1	HRA2_HUMAN
22	412.5	17.5	476	1	HRA4_HUMAN
23	412	17.5	400	1	YXAX_BACSU
24	410	17.4	458	1	HRA2_MOUSE
25	388	16.5	408	1	HTRA_LACLA
26	353	15.0	460	1	HRA3_MOUSE
27	344.5	14.6	321	1	HROA_ARATH
28	303.5	12.9	452	1	HRA3_HUMAN
29	230	9.8	630	1	Y4BJ_RHISN
30	198.5	8.4	997	1	YNN3_YEAST
31	137.5	5.8	767	1	DLG4_HUMAN
32	131	5.6	724	1	DLG4_RAT
33	130	5.5	724	1	DLG4_MOUSE

34	128	5.4	2021	1	OMPA_RICCN	Q52657 rickettsia
35	118	5.0	280	1	ETA_STAAN	P09331 staphylococ
36	115.5	4.9	837	1	APBL_HUMAN	Q02410 homo sapien
37	115	4.9	568	1	YELL_DROMD	Q09p71 drosophila
38	115	4.9	568	1	YELL_DROSU	Q02437 drosophila
39	114.5	4.9	931	1	DLGI_RAT	Q62696 rattus norv
40	114	4.8	904	1	DLGI_HUMAN	Q12959 homo sapien
41	114	4.8	2249	1	OMPA_RICRI	P15921 rickettsia
42	113.5	4.8	2364	1	PGCA_BOVIN	P13608 bos taurus
43	113	4.8	545	1	CH60_CAMJE	O69289 campylobact
44	111	4.7	261	1	DER3_DERPT	P39675 dermatophag
45	110.5	4.7	588	1	RUBB_BRANA	P21241 brassica na

ALIGNMENTS

RESULT 1  
DEGP\_RHIME  
ID DEGP\_RHIME STANDARD; PRT; 504 AA.  
AC Q52894;  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DE 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Probable serine protease do-like precursor (EC 3.4.21.-).  
GN DEGP1 OR DEGP OR R01021 OR SMC02365.  
OS Rhizobium meliloti (Sinorhizobium meliloti).  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Rhizobiaceae; Sinorhizobium.  
OX NCBI\_TaxID=382;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=96146524; PubMed=8550509;  
RA Glazebrook J., Ichige A., Walker G.C.;  
RT "Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in identification of degp: two loci required for symbiosis are closely linked to degp.";  
RL J. Bacteriol. 178:745-752(1996).  
[2]  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1021;  
RX MEDLINE=21396507; PubMed=11481430;  
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J., Boistard P., Becker A., Bounry M., Cadieu E., Dreano S., Gloux S., Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D., Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U., Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;  
RT "Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021.";  
Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
RL -1- SUBCELLULAR LOCATION: Periplasmic (Potential).  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DEGP/DEGP/DEGS FAMILY.  
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U31512; AAC43669.1; ALT\_INIT.  
CC EMBL; AL591785; CAC45593.1; -.  
CC MEROPS: S01.273; -.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR001940; Protease2C.  
CC InterPro; IPR000126; Ser\_proteas\_V8.  
CC InterPro; IPR001254; Trypsin.  
CC Pfam; PF00595; PDZ; 2.  
CC Pfam; PF00089; trypsin; 1.





```
Db 302 LGVQIQPTVKELSDSIGLKEAKGALITDPLKG-PAKAGIKAGDVIIISVNGEKINDVRDL 360
Qy 361 PVMVGAIITPGREVLGVRKGEIITIKAKLGNAAEHTGASSKTDEAPYT-EQSGTFSVE 419
Db 361 AKRIANMSPGETVITLGVWKSKEENIKVKLDSMPED---ENMKDGSKYSNEHCNSDETLE 417
Qy 420 SAGITLQTHTDSSGKHLVVRV---SDAERAGLRHGDILAV 459
Db 418 DYGLIVAPSDGCLG--LVVTDVDPDSDAADK-GIRPGDVIVTV 457

RESULT 3
DEGP_BRUAB
ID DEGP_BRUAB STANDARD; PRT: 513 AA.
AC Q44597;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=2308;
RX MEDLINE=95165990; PubMed=7861951;
RA Tatsum F.M., Cheville N.F., Morfitt D.;
RT "Cloning, characterization and construction of htra and htra-like
RL mutants of Brucella abortus and their survival in BALB/c mice.";
RL Microb. Pathog. 17:23-36(1994).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGQ/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U07352; AAA70164.1; -
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR Pfam; PF001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Signal; Periplasmic.
FT SIGNAL 1 25
FT CRAIN 26 513
FT DOMAIN 125 299
FT DOMAIN 300 391
FT DOMAIN 414 500
FT ACT_SITE 152 152
FT ACT_SITE 182 182
FT ACT_SITE 257 257
SQ SEQUENCE 513 AA; 53483 MW; DE1CEFL959472806 CRC64;

Query Match
Best Local Similarity 29.4%; Score 692.5; DB 1; Length 513;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy 9 AALACRALAGCEKAGSFFGADKKKESFVERIETHKDDGSYSMLLP---DFAQLVOSG 65
Db 14 AVALSRAAL-----AGAFVVTGPIGALNEARAE-----AVHVTTPPQAGFADLVEK 61
```

```
QY 66 PAVNIIQAAPAPRTONGSGN---AETSDPLADSPDPFFEFF-----KRLV 107
Db 62 PAVSVVRVKDVQETSNRGPQFGPGPDLPDGHPLKRFPRDFGMEPRGDSRSDNRCK 121
QY 108 PNMEIPEQERADDGGLNFGSGFIISKNGIYILTNTHVHVMAGMSIKVLLNDKREYTKLIGS 167
Db 122 ANKPRPGHERP-----VAQSGFVISEDGYVVTNNHVSVDGDAYTVLDDGTDLAKLGA 177
QY 168 DVOSDVALLKIDATE-ELPVVKIGNPKNLKPGEWAAIAGPFGDONSVTAGTIVSAKGRSL 226
Db 178 DPTDLAVLKLINAPKRFVYVAFGDNDKNRVGDWVAVGNPFLGGTGTSGIVSARGDI 237
QY 227 PNEYTPFIQTDVAINPFGSGPLFNKLGQVVGINSQIYSRSGGFMGIFSFAIPIDVAMV 286
Db 238 GAGPYDDFIQIDAAVNKGSGGPAFDLSGEVIGINTAIFSPGSGSVGIAFAIPSTAKOV 297
QY 287 AEQLKNTGKVQRGQLGVIIQVSVYGLAQSGFLDKRAGSALIAKILPGSPAERAGLQAGDIV 346
Db 298 VDQLIKKGSVERGWIGVQIQPVTKDIAASGLAAEKGAVASPDQDGPAAKAGIKAGDVI 357
QY 347 LSLDGGIRSGGDLPMVMVGAIITPGKEVSLGVWR--KGEEITIK-AKLGNAAEHTGASSKT 403
Db 358 TAVNGETVQDPRDLARKVANITAPGEKAALTVMRKNKABEINVTIAMPNDKSGSQSND 417
QY 404 DEAPYTEQSQGTFSVESAGITLQTHTDSSGKHLVVRV---SDAERAGLRHGDILAV 459
Db 418 NDGSGQE-----TLDVSLGTVVPSD--GKGVVTVDPDSDAADR-GIRSGDVIVSV 467

RESULT 4
DEGP_CHLMU
ID DEGP_CHLMU STANDARD; PRT: 497 AA.
AC Q9PL97;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
DE DEGP OR HTRA OR TC0210.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=MoPn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RL pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGQ/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AS002288; AAF39082.1; -
DR TIGR; TC0210; -
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; PR000126; Ser.proteas_v8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
```



```
Db 254 INGOVICNTAIVSGSGYIGIGFAISLMKRVIDQLISDQVTRGFLGTQLQIDSEL 313
Qy 313 AQSGFLDKASGLAKIATLPGSPAERAGLOAGDIVLSLDGIRSSGDLPLVMVGAITPGKE 372
Db 314 ATCYKLEKVGALVTDVVKGSPAERAGLQEDIVVAYNGKEVESLSALRNALISLMPGTR 373
Qy 373 VSLGVWRKGEITIKAKLGNAAHTGASSKTDEAPYTEQ-----SGTFSVESAGITLQTH 428
Db 374 VLKIVREG-----KTIETPVTVTQIPTEDGVSAQKMGVRVQNI 413
Qy 429 T-----DSSGHLVVRVSDAERAGLRHGDEILAV 459
Db 414 TPEICKLGLAADTRGILVAVEAGSPAASGAVPGQLILAV 455

RESULT 6
ID DEGP_CHLPN STANDARD; PRT; 488 AA.
AC Q9Z6T0; Q9J0D7; Q9KIW4;
DE 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR CPN0979 OR CP0877;
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWLO29;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RA "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RT Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.G.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RA "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWLO29 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGP/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001678; AAD19116.1;
DR EMBL; AE002246; AAF38665.1;
DR EMBL; AP002548; BAA99186.1;
DR MEROPS; S01.273; -.
```

```
DR PHCI-2DPAGE; Q9Z6T0; -.
DR TIGR; CP0877; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin..
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 20
FT CHAIN 21 488
FT DOMAIN 119 280
FT DOMAIN 281 372
FT DOMAIN 388 476
FT ACT_SITE 134 134
FT ACT_SITE 164 164
FT ACT_SITE 238 238
SQ SEQUENCE 488 AA; 52311 MW; 0EE7E0F8F106F49 CRC64;

Query Match 28.4%; Score 668.5; DB 1; Length 488;
Best Local Similarity 36.1%; Pred. No. 1.3e-35;
Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

Qy 10 LAALCAALLAGCEKAGSFFGADKKKAEFVERIETHTKDDGSVSMLLPDPFAQLVQSEGPVV 69
Db 10 LAVLVGSSLLALPLSGQAVG--RKES----RVSELPQDVLKKEISGGFSKVATKATPAVV 63

Qy 70 NQAAPAPR--TONGSNAETSDPLADSDPFY-EFFKRL--VPNMPPEIQ-EEADGGGL 123
Db 64 YIESFKSQAVTHPSFGRRGYPENPF---DYFNDEFFNRFFGLPSQREKPKQSKAVR--- 117

Qy 124 NFGSGFIISKNGYILNTHVHVGMSITKVLNDRKREYTKALIGSDVSDVALLKTDATDEE 183
Db 118 --GTGELVSPDGIYIVNNHVVEDTKIHLVLDHQQKTPATVIGLDPKTDLAVIKI-KSQN 174

Qy 184 LPVVKIGNPKNLKPGEMVAAGAPGFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAIN 242
Db 175 LPVLFGNSDHLKVGWDWAIAGNPFGLQATVTVGVISAKGRNQLRIADFEDFIOTDAAIN 234

Qy 243 PGNSGGLFLNLKQGVGINSQIYRSRSGFGMGISFAIPIDVAMNVAEOLKNTGKVQVGOLG 302
Db 235 PGNSGGLPLNIDGQVIGVNTAIVSGSGYIGIGFAISLMANRIIDQLIRDGQVTRGFLG 294

Qy 303 VIIEQVSYGLAQSGFLDKASGLIATLPGSPAERAGLOAGDIVLSLDGIRSSGDLPLV 362
Db 295 VTIQPIDAELACYKLEKVGALVTDVVKGSPAERAGLQEDIVVAYNGKEVDSLSMFRN 354

Qy 363 MVGAITPGKEVSLGVWRKGEITIKAKLGNAAHTGASSKTDEAPYTEQSGTFSVESAG 422
Db 355 AVSLMNPOTRIVLKVVRREGKVIETPVTVSQAPKEDGMS-----ALQRVG 398

Qy 423 ITLQTHDSSGKHL-----VVVRVSDAERAGLRHGDEILAV 459
Db 399 IRVQNLTPETAKKLGITAPETKGLIITSVEPGSAASGAVPGQLILAV 446

RESULT 7
HTOA_HAEIN STANDARD; PRT; 466 AA.
AC P45129;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable periplasmic serine protease do/hhoA-like precursor
DE (EC 3.4.21.-).
GN Hfl259.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
```

RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).  
CC -!- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA  
CC (PROTEASE DO) AND HHOA.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
CC DEGP/DECO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32805; AAC22906.1; -;  
DR MEROPS; S01.274; -;  
DR TIGR; H11259; -;  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Trypsin.  
DR Pfam; PF00595; PDZ; 2.  
DR Pfam; PF00089; trypsin; 1.  
DR PRINTS; PR00834; PROTEASES2C.  
DR SMART; SM00228; PDZ; 2.  
DR PROSITE; PS0106; PDZ; 2.  
KW Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.  
FT SIGNAL 1 29 POTENTIAL.  
FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE  
FT DO/HHOA-LIKE.  
FT DOMAIN 270 361 PDZ 1.  
FT ACT\_SITE 367 458 PDZ 2.  
FT ACT\_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).  
FT SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;  
Query Match 27.5%; Score 648; DB 1; Length 466;  
Best Local Similarity 37.6%; Pred. No. 2.4e-34;  
Matches 166; Conservative 76; Mismatches 163; Indels 36; Gaps 13;  
QY 29 GADKKKEASEVERIEHTKDDGVSMLLPDFQAQLVQSEGPVAVNIQAAPAPRTQNGSNAET 88  
DB 14 GLSLVSTSEVAHV-----AQATLPSEVFSEQNSLAPMLEKVQ--PAVVTLSVEGRKAV 63  
QY 89 DS-DPLADSDP--FYDFP-KRLVNNPPEIQEADGGGLNFGSGFII-SKNGYILTWTHV 143  
DB 64 DRSRPFLLDPIPEKFFKFFGDFAEAFQPGGRGSKRNFRGL--GSGVIINASKGYVLTNNHV 121  
QY 144 VAGMSGIKVLLDKREYTKALIGSDVQSVALLKIDATEELPVVYKIGNPNKLNKPGEWAA 203  
DB 122 IGAGKITVQLQDGRFEKAKLVGKQSDIALVQLKPSNLTEIKFADSKLRVGDFTVA 181  
QY 204 IGAPGFDSNVTAGIVSAKRSLPNES--YTPFIQTDVAINPGNSGGPLFNKLGQVVGIN 261  
DB 182 IGNPFLGLQTVTSIGIVSALGRSTGSDGTENYVYQTDAAVNRCSGGLVNLNGLIGIN 241

QY 262 SQIYSRSGFMGISFAIPIDVAMNVAEQKNTGKVQRGQLGVIQIEVSYGLAUSFGLDKA 321  
DB 242 TAIISPSGSGNAGIAFAIPSNQASNLVQILFQGVRRGLGKGGELNADAKAFNVSAQ 301  
QY 322 SGALIAKILPGSPAERAGLQAGDIVLSLOGGEIRSSGDLPMVMVGATTPGKEVSLGVWRKG 381  
DB 302 QGAFVSEVLLPKSAEKAGLKAGDIITAMNGOKITSSFAEIRAKITATTGACKEISLTYLRDG 361  
QY 382 EEITIKAKLGNAAHTGASSKTDDEAPYTEOOSGCTFSVESAGITLQTHTDSSGGHVLVVVR 441  
DB 362 KSHDVYMKL-QADSSQLSSSKT-ELPALD-----GATLKDYDAKGVKGIEITKI 408  
QY 442 ---SDAAERAGLRHGDILAV 459  
DB 409 QPNSLAAQR-GLKSGDIIIGI 428  
RESULT 8  
DEGP\_SALTY STANDARD; PRT; 475 AA.  
AC P26982;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Protease do precursor [EC 3.4.21.-].  
GN DEGP OR HTRA OR PTD OR STM0209.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C5;  
RX MEDLINE=91251770; PubMed=1645840;  
RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,  
RA Ali T., Miller I., Hormaeche C.;  
RT "The role of a stress-response protein in Salmonella typhimurium  
RT virulence.";  
RL Mol. Microbiol. 5:401-407(1991).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latrelle P.,  
RA Courtney L., Potwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
CC -!- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.  
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED  
CC SPECIFICITY WITH HHOA/DEGO.  
CC -!- SUBCELLULAR LOCATION: Periplasmic.  
CC -!- INDUCTION: BY HEAT SHOCK.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
CC DEGP/DEGO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X54548; CAA38420.1; -;  
DR EMBL; AE008704; AAL19173.1; -;  
DR PIR; S15337; S15337.  
DR PIR; S21327; S21327.



```
QY 173 VALLKIDATEELPVVKIGNPKNLKPGWVAAGAPFGDNSVTAGIYSAKGRS-LPNESY 231
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 140 IALLQIQNSKLTQATADSKLRVGFVAVAGNPGFGLGQTATSGIYSALGRSLNLEGL 199
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 232 TPIQTDVAINPNSGGLPFLNLKQVVGINSQIYSRSGGFMGISEFAIPIDVAMNVAEOLK 291
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 ENFIQTDASINRSGSALLNLGELIGINTAILAPGGSGVIGFAIPSNMARTLAQOLI 259
:||||: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 292 NTCKVORGOLGVIOEVSYGIAQSGFLDKASGALIAKTLPGSPAERAGLOAGDIVLSLDG 351
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 DFEIKRGLGIGKGTENSADIAKAFNDVORGAEVSEVLPGSGSAKAGVKGADIIISUNG 319
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 352 GEIRSSGDLPMVVGAIITPGKSEVSLGVWRKGEITIKAKLGNAAEHTGASSKTDEAPYTEQ 411
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 320 KPLNSFAELRSRIATTEPGKVKVGLLNRKNGKPLEVEVTLDTST----- 362
QY 412 OSTFVSAGITLQHTDSSG-----KHLVVVVSDAARAGLRHGHDEILAV 459
: : : : : : : : : : : : : : : : : : : : : : : : : :
363 -SSASAEMITPALLEGATLSDGOLKDGKGKIKIDEVYKGSPPAQ-AGLQKDDVIIGV 417

RESULT 10
DEGP_RICPR STANDARD; PRT; 513 AA.
AC 005942;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR RPI24.
OS Rickettsia prowazekii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=782;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MDR12 E;
RX MEDLINE=99039499; PubMed=9823893;
RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
RA Sacheritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
RA Eriksson A.-S., Winkler H.H., Kurland C.G.:
RT "The genome sequence of Rickettsia prowazekii and the origin of
RT mitochondria.";
RL Nature 396:133-140(1998).
RN [2]
RP SEQUENCE OF 161-513 FROM N.A.
RC STRAIN=MDR12 E;
RX MEDLINE=97419517; PubMed=9274032;
RA Andersson J.O., Andersson S.G.E.;
RT "Genomic rearrangements during evolution of the obligate intracellular
RT parasite Rickettsia prowazekii as inferred from an analysis of 52015
RT bp nucleotide sequence.";
RL Microbiology 143:2783-2795(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGQ/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ235270; CAA14593.1; -
DR EMBL: Y11782; CAA72471.1; -
DR MEROPS: S01.273; -
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
```

```
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 513
FT DOMAIN 124 289
FT DOMAIN 290 381
FT DOMAIN 418 502
FT ACT_SITE 139 139
FT ACT_SITE 169 169
FT ACT_SITE 247 247
SQ SEQUENCE 513 AA; 56309 MW; 4B7E9B7AB4079139 CRC64;

Query Match 26.2%; Score 615.5; DB 1; Length 513;
Best Local Similarity 35.4%; Pred. No. 3.2e-32;
Matches 149; Conservative 84; Mismatches 157; Indels 31; Gaps 10;

QY 57 FAOLYQSEGPVAVNIOAAPAPRTQNGSNAETSDPLADSDPEYEPF----KRLVP-NM 110
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 56 FADIVEPLIPAVVNI--STIEYVNDKSENSEK- ---LLOENKHLGFMFSDVLEKLNIPNL 110
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 111 PEIQEEADDDGLNFGSGFIISKNGYILTNTHVYVAGMSGIKVLLNDKREYTAKLIGSDVQ 170
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 EETAKTPK---SIPLGSGFIIAPNGLIVTNHYHVIANVEKINIKLADNTEFLAKLIGSDSK 167
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 171 SDVALLKIDATEELPVVKIGNPKNLKPGWVAAGAPFG-FDINSVTAGIYSAKGRSLPNE 229
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 168 TDLALLKIDSEELPLPVEFGSDNDARVGDWVIAIGNPGFGLGVTGVIISSKGRDIDVD 227
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 230 S---YTPFTQTDVAINPNSGGLPFLNLKQVVGINSQIYSRSGGFMGISEFAIPIDVAMNV 286
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 TDNIVDNFIQTDAAINNGSGGPMFNLDQKVGICVNTAIFSPCLCTNIGIGFAIPNTAKPI 287
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 AEOLKNTKVVORGOLGVIOEVSYGIAQSGFLDKASGALIAKTLPGSPAERAGLOAGDIV 346
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 288 IERLKKDQKVSRLGVTIQDLTEIESEVLGFKGTNGVLVSKVQENGPGYKAGIKKKGDII 347
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 347 LSLDGEIRSSGDLPMVVGAIITPGKSEVSLGVWRKGE--EITIKAKLGNAAEHTGASSKTD 404
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 IKFGRLVKNYTKKLRVVIADTPINQEVKLIKLRDAQELPLIKVTADNEEVINDSTEETN 407
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 405 EAPYTEQSGSTFVSAGITLQHTDSSGKH-----LVVVSYSDAARAGLRHGHDE 455
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 408 KAVILINKKENLSITKNNTFSNLTEELRKKYDIPQDKTGIVINIIDE--EESVFKLGDL 465
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 456 I 456
Db 466 I 466

RESULT 11
DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR 20173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia
```





```
Db 38 MPSLAPMLEKVPMSVSVINV-----EGSTTVNTPRMPRNFOQFFGDDSPFCQEGSPF- 89
QY 108 PNMPEIQEADDDG-----LNFSGSFI-SKNGYILTTHVVAGMSIKVLNDKREY 160
Db 90 QSSPFCGGGGGQGGQKFMALGSGVIIDADKGVYVTVNHHVVDNATVIKQLSDGRKF 149
QY 161 TAKLIGSDVQSDVALLKIDATEELPVVKIGNPNLKPGEWVAIGAPFGFSDNSVTAGIVS 220
Db 150 DAKMVGKDRPSDIALIQIQNPKNLTAIKMADSDALRVGDTVAIGNPFGICETVTSIGVS 209
QY 221 AKGRS-LPNESTYPTFOTDVAIINPGNSGGPLENLKGOVGVINSQIYSRSGFGWGISFAIP 279
Db 210 ALGRSGLNENYENFTQDAIINRSGSGGALVNLNGELIGINTAILAPDGGNIGIGFAIP 269
QY 280 IDVMNVAEQLKNTGKVQSGQGVIIQEVSYGLAQSGFLDKASGALIAKILPQSPABRAG 339
Db 270 SNMKNLTQMVVEYGVQKRGELGIMCTELNSELAKAMKVDAQRGAFAVSQVLPNSSAAKAG 329
QY 340 LQAGDIVLSLGDGEIRSSGDLPVWVGAIPTGKEVSLGVWRKGEIITKAKLGNAAEHTGA 399
Db 330 IKAGDVITSLNGKPISSFAALRAQVQTMPVGVSKLTGLLRDQGVNVNLEL----- 380
QY 400 SSKTDEAPYTEQSGGFFVESAE-----GITLQHTDSSGKHLVVVVRVSDAAERAGLRH 452
Db 381 -----QSSQSQVDSISFNGIEGAEMSKGKQGVVNVNVTGTPTAAQIGLKK 429
QY 453 GDEIL 457
Db 430 GDVII 434

RESULT 12
DEGP_BUCAP
ID DEGP_BUCAP STANDARD; PRT; 478 AA.
AC P57322;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR BU228.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OC symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
MDLINE=20445173; PubMed=10993077;
Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.:
"Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB12943.1; -.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001940; Protease2C.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 2.
CC PROSITE; PS50106; PDZ; 1.
```

```
KW Hydrolase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 478
FT DOMAIN 116 254
FT DOMAIN 281 372
FT DOMAIN 387 469
FT ACT_SITE 133 133
FT ACT_SITE 163 163
FT ACT_SITE 238 238
SQ SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;

Query Match 25.3%; Score 594.5; DB 1; Length 478;
Best Local Similarity 34.1%; Pred. No. 6.5e-31;
Matches 150; Conservative 86; Mismatches 151; Indels 53; Gaps 10;

QY 43 HTKDDGVSMLLPDPAQLVQSEGPVAVVNIQAAPAPRTQNGSNAETSDPL-----A 94
Db 31 NTKNSIVSREISPSLAPMLEKVMPSVISI-----NIEGSAITRTSRLPHQFPFFG 81
QY 95 DSDPF-----YEFKRLVPNMPPEIPOEADDGGLNF---GSGFIISKN-CYILTNT 142
Db 82 DNSPFCQGNPSFRHSPFCHINP-----DSDDKKEKFRALGSGVLIINADKCYAVTN 133
QY 143 VVAGMSIKVLNDKREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPNLKPGEWVA 202
Db 134 VVENANKIQVQLSDGRRYEARVIGKDRSDIALIQLKNNANLSEIKIADSDNLRVGDYTV 193
QY 203 AIGAPFGFSDNSVTAGIVSAKGRS-LPNESTYPTFOTDVAIINPGNSGGPLENLKGOV 261
Db 194 AIGNPYLGIVTSGIISALGRSGLNIEHYENFIQTDAAINRSGSGGALVNLNGELIGIN 253
QY 262 SQIYSRSGFGMGISPAIPIDVAMNVABEOLKNTGKVQSGQGVIIQEVSYGLAQSGFLDKA 321
Db 254 TAILAPDGGNIGIGFAIPCNVKNLTQMVQGVQVRRGELGIMGHELNDSLAQIMKINSQ 313
QY 322 SGALIAKILPQSPABRAGLQAGDIVLSLGDGEIRSSGDLPVWVGAIPTGKEVSLGVWRK 381
Db 314 KGAFVSRVLPNSSAFAEAGIKAGDIIISLNRKPISSFSLSRAETGSLPVATKMGVFRG 373
QY 382 EEITKAKLGNAAEHTGASSTDEAPYTEQSGTFSVESAGITLQHTDSSGKHLVV--V 439
Db 374 RKNITVELKHSVKH-NLNSND-----YIGIEGVDLSYIFNEOKVTKVDNV 420
QY 440 RVSDAAERAGLRHGDILAV 459
Db 421 KPHTPASKIGFKKDDIILNV 440

RESULT 13
DEGP_BUCAP
ID DEGP_BUCAP STANDARD; PRT; 478 AA.
AC O85291;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98353428; PubMed=9688822;
RA Thao M.L., Baumann P.:
"Sequence analysis of a DNA fragment from Buchnera aphidicola (Aphid
endosymbiont) containing the genes dapD-htra-ilvi-ilvh-ftsl-ftsi-
murE.";
RL Curr. Microbiol. 37:214-216(1998).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC
```



This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announcement/or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announcement/or_send_an_email_to_license@isb-sib.ch)).

EMBL:	AF060492;	AAC32331.1;	-
DR	MEROFS;	S01.273;	-
DR	InterPro:	IPR001478;	PDZ.
DR	InterPro:	IPR001940;	Protease2C.
DR	InterPro:	IPR001254;	Trypsin.
DR	Pfam:	PF00595;	PDZ; 2.
DR	Pfam:	PF00089;	trypsin; 1.
DR	PRINTS:	PR00834;	PROTEASES2C.
DR	SMART:	SM00228;	PDZ; 2.
DR	PROSITE:	PS50106;	PDZ; 1.
DR	Hydrolase;	Serine protease;	Signal.
KW	SIGNAL	1	POTENTIAL.
FT	CHAIN	27	478
FT	DOMAIN	116	254
FT	DOMAIN	281	372
FT	DOMAIN	387	469
FT	ACT_SITE	133	133
FT	ACT_SITE	163	163
FT	ACT_SITE	238	238
FT	SEQUENCE	478 AA;	51303 MW; C044824F7EF4E98E CRC64;
DR	PROBLEME SERINE PROTEASE DO-LIKE.		
DR	CATALYTIC.		
DR	PDZ 1.		
DR	PDZ 2.		
DR	CHARGE RELAY SYSTEM (POTENTIAL).		
DR	CHARGE RELAY SYSTEM (POTENTIAL).		
DR	CHARGE RELAY SYSTEM (POTENTIAL).		
DR	CHARGE RELAY SYSTEM (POTENTIAL).		

Query Match 24.3%; Score 571.5; DB 1; Length 478;  
Best Local Similarity 32.4%; Pred. No. 1.9e-29;  
Matches 142; Conservative 81; Mismatches 146; Indels 69; Gaps 9;

QY	53	LLPDFAQILVQSEPAV--VNIQAAPRTON-----GSGNAETSDSLAD	95
		: : :   :   :   :   :   :   :	
Db	41	LAPSLAPMLEKVPMSVISTNIEGTSVVVHTSRLPHQHPQFFGHNSPFCQGNSPFRNSPFCR	100
QY	96	SDP----FYEFFKRLVPNMPEIQEEADGGGLNFGSGFIISKN-GYTLTHTVVAGMSI	150
Db	101	:     :   :   :   :   :   :   :   :   :	
		:     :   :   :   :   :   :   :   :   :	
QY	151	KVLLNDKREYTAKLQSDVQSDVALLKIDATEELPVVYKIGNKPNLKPGEWVAAGAPFGF	210
Db	142	QVQLSDGRRYEASTICKDSDRIALIQLKNAKLSAIKIADSDTLRVGDVTVAIGNPVGL	201
QY	211	DNSVTAGIVSAKRS-LPNESYTPFTQTDVAINPGNSGGPLFNLKGQVGINSOITRSRG	269
Db	202	GETVSGIITSALRGSGNIEHYENFIQTDAAINRGNSGGALVNLKGLIGINTAILAPDG	261
		:   :   :   :   :   :   :   :   :   :   :	
QY	270	GPWGISFAIPIDVAMNVAEOLKNTGKVORGOLGVITQEVSYGLAQSPGLBKASCALIARI	329
Db	262	GNIGTGAIPGNVKNLTQOMVFGQVKRGELIGIMELNSDLAHVMKINAQGAFFVSQV	321
QY	330	LPGSPAERAGLQAGDILVLDGGEIRSSGDLPMVMGAITPGKEVSLGKWRKGE-----IT	385
Db	322	LPNSAFHAGIKAGDIIVSLNKKTISSFAALRAEVSGLPVSTRKMLGIFRNGITKNVIVE	381
QY	386	IRAKLGNAB---HTGASKTDEAPYTEOQSSTFVESAGITLQHTDSSGKHLVVVRV	441
Db	382	LPKLSKNSVSLGDITYTG-----IEGADLSDCSLNGOKGVKIEIKL	422
QY	442	SDAAERAGLRHGDEILAV	459
Db	423	NTQASKIGPKDDIIVEV	440
		:   :   :   :   :   :   :   :   :   :   :	
		:   :   :   :   :   :   :   :   :   :   :	
RESULT	14		
DEGL ARATH			
ID	DEGL ARATH	STANDARD;	PRT; 437 AA.
AC	Q22609; Q9ULK85;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		

FT CONFLICT 36 36 V -> I (IN REF. 2).  
FT CONFLICT 54 54 P -> S (IN REF. 2).  
FT CONFLICT 60 60 G -> R (IN REF. 2).  
FT CONFLICT 64 64 G -> D (IN REF. 2).  
FT CONFLICT 68 69 LL -> HF (IN REF. 2).  
FT CONFLICT 355 355 L -> V (IN REF. 2).  
FT CONFLICT 381 381 I -> V (IN REF. 2).  
FT CONFLICT 416 416 Q -> E (IN REF. 2).  
SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;

Query Match 20.0%; Score 470; DB 1; Length 437;  
Best Local Similarity 30.3%; Pred. No. 5e-23;  
Matches 139; Conservative 70; Mismatches 167; Indels 82; Gaps 13;

QY 9 ALAALCAALL-----ACCEKAGKFFGADKKEASFVE-----RI-----EHTK 45  
DB 2 AITTSCLLLHSPPSQSLNSISFFNLSSRSFVSLPIRSKRYFRILSKPLPLNDNGD 61  
46 DGSVSMLLP-----DFAQLVSEGPAVVNIQAAPRTQNG---SGNAETSDPLA--- 94  
DB 62 DGDGDTLLLPFSAVKPFELLCISVALSFLFAASPAVESASAFVWSTPKKLQTDDELATVR 121  
QY 95 ----DSDPFYEFKRLVP-----NNPEIQEADGGNFGSGFIISKNGYILTNTHV 144  
DB 122 LQENTPSVYITNLAVRODAFTLDVLEVPQ-----GSGSGFVMDKQGHIVTNHVI 173  
QY 145 AGMSGIKVLNDKREYTAKLIGSDVQSVALLKIDATE-ELPVVKIGKNPKNLKPGEWAA 203  
DB 174 RCASDLRVTLAQDTFDKAVGFDQDKDVAVLRIAPKKNLRIPYGVSAADLLVGKQVFA 233  
QY 204 IGAPGFDSNVTAGIVSAKGRSLPNES-----YTPFTOTDVAINPGNSGGFLNKGQVVG 259  
DB 234 IGNPGLDHTLTGTVTSGLRREISSAATGRPIQDVITQDAAINPGNSGGFLDSSGTLIG 293  
QY 260 INSOIYRSRSGFNGISFAIPIDVAMVABOLKNTGKVGORQLGVITQEVSYGLAQSGLD 319  
DB 294 INTAISPAGSGGSGVGFSPVDVGVGIVDQVLRFGKVTIRPLG-----IKFAPDQSVLEQL 348  
QY 320 KASGALIAKILPGSPAERAGLQA-----GDIVLSLDGGEIRSSGDLPLVMVGAIT 368  
DB 349 GVSGLVLLDAPPSPGAGKAGLQSTKRDGVRGLILGDIITSVNCTKVSNGSDLYRLDQCK 408  
QY 369 PKKEVSLGVWRKGEETIKAKIGNAAEHTGASSKTDEA 406  
DB 409 VGDEVTVQVLRGDHKEKISVTL-----EPKPDES 437

## SULT 15

ID HRAI\_HUMAN STANDARD; PRT; 480 AA.  
AC Q92743; Q9UN53;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Serine protease HTRA1 precursor (EC 3.4.21.-) (L56).  
GN PRS11 OR HTRA1 OR HTRA.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Placenta;  
RC MEDLINE=97131593; PubMed=8977104;  
RA Zimbrunn J., Trueb B.;  
RT "Primary structure of a putative serine protease specific for IGF-binding proteins.";  
RL FEBS Lett. 398:187-192(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Crowl R.M., Luk D., Milnamow M.;  
RT "Genomic organization and promoter characterization of the human HTRA

(PRS11) gene.";  
Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
[3]  
SEQUENCE OF 144-480 FROM N.A., AND MUTAGENESIS OF SER-328.  
TISSUE=Cartilage;  
RX MEDLINE=99069438; PubMed=9852107;  
RA Hu S.I., Carozza M., Klein M., Nantermet P., Luk D., Crowl R.M.;  
RT "Human HTRA, an evolutionarily conserved serine protease identified as a differentially expressed gene product in osteoarthritic cartilage.";  
J. Biol. Chem. 273:34406-34412(1998).  
CC -!- FUNCTION: Protease that regulate the availability of IGFs by cleaving IGF-binding proteins.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- TISSUE SPECIFICITY: Expressed in a variety of tissues, with strongest expression in placenta.  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE DSGP/DEGP/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 1 IGFBP DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y07921; CAA69226.1; -;  
DR EMBL; AF157623; AAD41525.1; -;  
DR EMBL; AF097709; AAC97211.1; -;  
DR HSP; P80424; 1AN1.  
DR MEROPS; S01.277; -;  
DR MIM; 602194; -;  
DR InterPro; IPR000867; IGFBP.  
DR InterPro; IPR001478; PDZ.  
DR InterPro; IPR001940; Protease2C.  
DR InterPro; IPR001254; Trypsin.  
DR InterPro; IPR002350; Kazal.  
DR Pfam; PF00219; IGFBP; 1.  
DR Pfam; PF00050; kazal; 1.  
DR Pfam; PF00595; PDZ; 1.  
DR PRINTS; PF00089; trypsin; 1.  
DR PRINTS; PF00834; PROTEASES2C.  
DR SMART; SM00121; IB; 1.  
DR SMART; SM00280; KAZAL; 1.  
DR SMART; SM00228; PDZ; 1.  
DR PROSITE; PS50106; PDZ; 1.  
KW Hydrolase; Serine protease; Growth factor binding; Signal.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 480 SERINE PROTEASE HTRA1.  
FT DOMAIN 37 94 IGFBP.  
FT DOMAIN 101 155 KAZAL-LIKE.  
FT DOMAIN 204 364 SERINE PROTEASE.  
FT DOMAIN 365 467 PDZ.  
FT ACT\_SITE 220 220 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 250 250 CHARGE RELAY SYSTEM (POTENTIAL).  
FT ACT\_SITE 328 328 CHARGE RELAY SYSTEM.  
FT MUTAGEN 328 328 S->A: LOSS OF ACTIVITY.  
FT CONFLICT 323 323 I -> T (IN REF. 3).  
SQ SEQUENCE 480 AA; 51286 MW; CA20A99480FB2330 CRC64;

Query Match 19.0%; Score 448; DB 1; Length 480;  
Best Local Similarity 33.9%; Pred. No. 1.4e-21;  
Matches 118; Conservative 61; Mismatches 103; Indels 66; Gaps 10;

QY 58 AQLVQSEGPAVVNIQAAPRTQNGSGNAETSDPLADSPFYEFKRLVPMNPEIQOE 117  
DB 173 ADVVEKTIAPAVVHI-----ELFRKLPFSKREVP--- 200  
QY 118 ADDGGLNFGSGFIISKNGYILTNTHVAGMSGIKVLNDKREYTAKLIGSDVQSVALLK 177

Db 201 : |||||:|:| ||| :|| | : | || : | :|:|:| 255  
Qy 178 IDATEELPVVKIGNPKNLKPGEWAAIGAPFGFNDNSVTAGIYSAK---GRSLP-NESYTP 233  
Db 256 IDHOGKLPVLLGRSSSELRPGEFVVAIGSPFSLQNTVTGTIVSTTORGKELGLRNSDMD 315  
Qy 234 FIOTDVAINPGSGPLFNLKGQVVGINSQIYSRSGFGMGSFAIPID-----VAMNVAE 288  
Db 316 YIOTDAIINYNGSGPLVNLDEVEIGINTLKV-----AGISFAIPSDKIKKFLTESHDR 370  
Qy 289 QLKNTGKVRQOLGVIIQEVSYGLAQSGF-----LDRKASGALIAKILPGSPAERAGLOA 342  
Db 371 QAKGKAITKKYIGIRMMSLTSSRAKELDRHRDPDVISGAYIIIEVIPDTPAEAGGLKE 430  
Qy 343 GDIVLSLDGGEIRSSGDLPMVMVGATPGKEVSLG-VWRKGE---ITI 386  
Db 431 NDVIISINGQSVVSANDVSDVI-----KRESTLNMVVRGNEIDIMITV 473

Search completed: June 10, 2002, 12:14:14  
Job time: 223 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 10, 2002, 12:09:56 ; Search time 31.16 Seconds  
(without alignments)  
2581.600 Million cell updates/sec

Title: US-09-388-090-4  
Perfect score: 2353  
Sequence: 1 VFKKYQYFALALCAALLAG.....ERAGLRHGDEILAVRASPRQ 465

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: SPTEMBL19:\*
  - 2: sp\_archaea:\*
  - 3: sp\_bacteria:\*
  - 4: sp\_fungi:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phase:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvrius:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	16	Q9JVT1 neisseria m
2	989.5	42.1	474	2	Q9ALS1 pseudomonas
3	985.5	41.9	474	16	Q57155 pseudomonas
4	894.5	38.0	481	2	Q9AOD1 pseudomonas
5	862.5	36.7	514	16	Q9PBA3 xylella fas
6	746	31.7	513	16	Q98CS8 rhizobium l
7	743	31.6	473	2	Q44476 azotobacter
8	719.5	30.6	516	16	Q985F9 rhizobium l
9	673	28.6	504	16	Q98KJ1 rhizobium l
10	664.5	28.2	500	2	Q44652 brucella ab
11	660	28.0	433	2	Q9FDL1 aeromonas h
12	639.5	27.2	478	2	P74978 yersinia en
13	638	27.1	511	16	Q92JA1 rickettsia
14	637	27.1	437	2	O68197 haemophilus
15	637	27.1	459	16	Q9CMS7 pasteurella
16	637	27.1	474	2	Q44596 brucella ab

17	636	27.0	503	16	Q926C8 rhizobium m
18	636	27.0	530	16	Q9A4S2 caulobacter
19	632.5	26.9	456	16	Q9KUF5 vibrio chol
20	632	26.9	463	2	O68198 haemophilus
21	621.5	26.4	500	2	O9KJN6 myxococcus
22	614.5	26.1	459	16	Q9WZ41 thermotoga
23	606.5	25.8	452	16	P73354 synecocyst
24	606.5	25.8	491	2	Q9L8K0 shigella so
25	604.5	25.7	481	16	Q9PGL3 xylella fas
26	603.5	25.6	371	2	O31388 bradyrhizob
27	600.5	25.5	416	16	P73940 synecocyst
28	600	25.5	460	2	O06439 rhodobacter
29	589	25.0	389	16	Q9HVL1 pseudomonas
30	582	24.7	476	16	Q9ZM18 helicobacte
31	578	24.6	428	16	Q98N31 rhizobium l
32	575	24.4	464	2	Q56885 yersinia en
33	574.5	24.4	443	16	O25663 helicobacte
34	571	24.3	453	16	O67436 aquifex ae
35	545.5	23.2	394	16	P72780 synecocyst
36	545	23.2	472	16	Q9A8R9 caulobacter
37	544.5	23.1	465	16	Q92QE6 rhizobium m
38	529	22.5	466	2	Q53247 rickettsia
39	524	22.3	466	2	O05335 rickettsia
40	524	22.3	466	2	Q53246 rickettsia
41	522	22.2	466	2	Q53249 rickettsia
42	511.5	21.7	472	16	Q9PN69 campylobact
43	511	21.7	466	2	Q53251 rickettsia
44	509	21.6	389	2	Q51374 pseudomonas
45	504.5	21.4	472	2	Q46120 campylobact

ALIGNMENTS

RESULT	ID	Q9JVT1	PRELIMINARY;	PRT;	499 AA.
AC	Q9JVT1	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)			
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)			
DE	PUTATIVE PERIPLASMIC SERINE PROTEASE (EC 3.4.21.)				
GN	NMA0710				
OS	Neisseria meningitidis (serogroup A).				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=65699;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;				
RX	MEDLINE=20222556; PubMed=10761919;				
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltham R., Hamlin N., Holroyd S., Jagals K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrett B.G.;				
RT	"Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."				
RL	Nature 404:502-506(2000).				
CC	- I - SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.				
DR	EMBL: AL162754; CAB83996.1; -				
DR	MEPROS; S01UPC; -				
DR	InterPro: IPR001478; PDZ.				
DR	InterPro: IPR001940; Protease2C.				
DR	InterPro: IPR001254; Trypsin.				
DR	Pfam: PF00595; PDZ; 2.				
DR	Pfam: PF00089; trypsin; 1.				
DR	PRINTS: PR000834; PROTEASES2C.				
DR	SMART: SM00228; PDZ; 2.				
DR	PROSITE: PS0106; PDZ; 1.				
DR	PROSITE: PS0240; TRYPSIN_DOM; 1.				
KW	Complete proteome; Hydrolase; Protease; Serine protease.				
SEQ	SEQUENCE 499 AA; 52563 MW; 73E979F0A7F5B969 CRC64;				

```
Query Match          95.5%; Score 2246; DB 16; Length 499;
Best Local Similarity 95.9%; Pred. No. 6.1e-132;
Matches 444; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAALCAALLACERAGSFFGADKKKEASVERIEHTKDDGSVSMLLPDFQAQL 60
Db 1 MFKKYQYFALAALCAALLACERAGSFFGADKKKEASVERIEHTKDDGSVSMLLPDFVQL 60
QY 61 VOSEGPVAVNIQAAPRTONGSGNAETSDPLADSDPFYEFFKRLVNPMPPEIQEEADD 120
Db 61 VOSEGPVAVNIQAAPRTONGSGNAETSDPLADSDPFYEFFKRLVNPMPPEIQEEADD 120
QY 121 GGLNFGSGFIISKNGYILTNTHVAGMSIKVLLNDRKREYTAKLIGSDVQSDVALLKIDA 180
Db 121 GGLNFGSGFIISKNGYILTNTHVAGMSIKVLLNDRKREYTAKLIGSDVQSDVALLKIDA 180
QY 181 TEELPVVKGKGNPKLPGGEVAAIGAPFGFDSNVTAGIVSAKGRSLPNESTYPTIQTDDVA 240
Db 181 TEELPVVKGKGNPKLPGGEVAAIGAPFGFDSNVTAGIVSAKGRSLPNESTYPTIQTDDVA 240
QY 241 INPGNSGGPLNFKGVVINSOIYSRSGGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 300
Db 241 INPGNSGGPLNFKGVVINSOIYSRSGGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 300
QY 301 LGVIIQEVSYGLAQSEGLDKAGGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360
Db 301 LGVIIQEVSYGLAQSEGLDKAGGALIAKILPGSPAERAGLQAGDIVLSLDGGEIRSSGDL 360
QY 361 PVMVGAIITPGKESVLSGWRKGEITITKALGNAAEHTGASSKTDEAPYTEOQSSTFSVES 420
Db 361 PVMVGAIITPGKESVLSGWRKGEITITKALGNAAEHTGASSKTDEAPYTEOQSSTFSVES 420
QY 421 AGITLTHTDSSGKHLVVVRSVDAEERAGLRHGDIEILAVRAS 463
Db 421 AGITLTHTDSSGKHLVVVRSVDAEERAGLRHGDIEILAVRAS 463

RESULT 2
Q9ALS1 PRELIMINARY; PRT; 474 AA.
AC Q9ALS1
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD.
DE MUCD.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UCBPP-PA14;
RA Yorgey P.S., Rahme L.G., Tan M., Ausubel F.M.;
RT "The Roles of mucD and Alginate in the Virulence of Pseudomonas
aeruginosa in Plants, Nematodes, and Mice."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF343973; AAK11276.1;
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 474 AA; 50349 MW; A22FD4338B859D4C CRC64;
```

```
Query Match          42.1%; Score 989.5; DB 2; Length 474;
Best Local Similarity 52.5%; Pred. No. 9.1e-54;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTONGSGNAETSDPLADSDP-FYEFKRLVNP 109
Db 27 LPDFTPLVEQASPAVVNISTRKLP-----DRAMARQQLSIPDLEGUPPMFRDFLERSIPQ 82
QY 110 MFEIP---QEEADGGLNFGSGFIISKNGYILTNTHVAGMSIKVLLNDRKREYTAKLIG 166
Db 83 VPRNPRGQREAO-----SLGSGFIISNDGYILTNHVVADAEILVRLSDRSEHKAKLVG 138
QY 167 SDVQSDVALLKIDATEELPVVKGKGNPKLPGGEVAAIGAPFGFDSNVTAGIVSAKGRSL 226
Db 139 ADPRSDVAVLKIEA-KNLPTLKLGDGNKLVGEMWLAIGSPFGFDHSTVAGIVSAKGRSL 197
QY 227 PNESTYPTIQTDDVAINPGNSGGPLNFKGVVINSOIYSRSGGFMGISFAIPIDVAMNV 286
Db 198 PNESTYPTIQTDDVAINPGNSGGPLNFKGVVINSOIYSRSGGFMGISFAIPIDVAMNV 257
QY 287 AEOLKNTCKVORGQVGVVINSOIYSRSGGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 346
Db 258 ADOLKNTCKVORGQVGVVINSOIYSRSGGFMGISFAIPIDVAMNVAEOLKNTCKVORGQ 317
QY 347 LSLDGEIRSSGDLPMVGAIITPGKESVLSGWRKGEITITKALGNAAEHTGASSKTDE- 405
Db 318 LSLNQSINESADLPHLVGNMKPGDKINLDVIRNCGKSLSMVAGNLPD-----DDEE 370
QY 406 ----APYTEOQSSTFSVESAGITLTQTH--TSSGKHLVVVRSVDAEERAGLRHGDIEI 456
Db 371 IASMGAPGAERSNRLGVTVADLTAEQRKSLDIQG-GVVIKEVODGPAVIGLRPGDVI 428

RESULT 3
Q57155 PRELIMINARY; PRT; 474 AA.
AC Q57155
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD (SERINE PROTEASE MUCD).
DE MUCD OR PA0766
CN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=93391358; PubMed=8378309;
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
RA Deretic V.;
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
infecting cystic fibrosis patients."
RN Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=95286510; PubMed=7768826;
RA Yu H., Schurr M.J., Deretic V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
aeruginosa AlgU: E. coli rpoE restores mucoidy and reduces sensitivity
to reactive oxygen intermediates in algU mutants of P. aeruginosa."
RL J. Bacteriol. 177:3259-3268(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=96134987; PubMed=8550474;
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
RA Deretic V.;
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
aeruginosa in cystic fibrosis encode homologs of the serine protease
HtrA.";
```

```
RL J. Bacteriol. 178:511-523(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; U49151; AAC43718.1; -.
DR EMBL; U32853; AAC43676.1; -.
DR EMBL; AE004511; AAG04155.1; -.
DR MEROPS; S01.UPC; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBE CRC64;

Query Match 41.9%; Score 985.5; DB 16; Length 474;
Best Local Similarity 52.5%; Pred. No. 1.6e-53;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTQNGSNAETSDPLADSDP--FYEFFKRLVP 109
DB 27 LPDFTPLVEQASPAVAVNISTRQKLP-----DRAMARGQLSIPDLEGLPPMFRDLEERSIPQ 82
QY 110 MPPIP---QSEADGGGLNFGSGGFTISKNGYILTTTHVVAGMGSIKVLNDRKREYTA 166
DB 83 VPRNPRCQREQA-----SLSGGFIISNDGYILTTNNHVVADADEILVRLSDRSEHKALIG 138
QY 167 SDVQSDVALLKIDATEELPVVKIGNPKNKPGEWVAAGAPFGFDSNVTAGIVSAKGRSL 226
DB 139 ADPRSDVAVLKIEA-KNLPTLKLGDNSNKLKVGWVLAIGSPFGFDSHVSVTAGIVSAKGRSL 197
QY 227 PNEYTPFTQTDVAINPGNSGGPLNKGQVVGINSQIYRSRSGFGMGISFAIPIDVAMNV 286
DB 198 PNEYTPFTQTDVAINPGNSGGPLNKGQVVGINSQIYRSRSGFGMGISFAIPIDVAMNV 257
QY 287 AEOLKNTGKVGQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAGDIV 346
DB 258 ADQLKAGKGVSRGWLGVVIOEVNKLDAESFGLDKPAGLVAQLVEDGPAAGGLQGVGDI 317
QY 347 LSLDGGEIRSSGDLPMVVGVAITPGKEVSLGVWRKGEETITKAKLGNAAEHTGASKTDE- 405
DB 318 LSLNGOSINESADLPHLVGNMKGDKINLDVIRNGQRKSLSMAYGSLPD-----DDEE 370
QY 406 -----APYTEQQSGTFSVESAGITLQTH--TDSSGKHLVVVVSDA-AERAGLRHGEI 456
DB 371 IASMGAPGAERSNRNLGVTVAADLTAEQKKSLLDIQG-GVVIKEVQDGPAAVIGLRPGDVI 428

RESULT 4
Q9AQD1 PRELIMINARY; PRT; 481 AA.
AC Q9AQD1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MUCD.

GN MUCD.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FF5;
RA Keith L.M.W., Bender C.L.;
RT biosynthesis and response to environmental stress in Pseudomonas
RT syringae.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF190580; AAK01318.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 481 AA; 50601 MW; E2C9C11137B83920 CRC64;

Query Match 38.0%; Score 894.5; DB 2; Length 481;
Best Local Similarity 47.6%; Pred. No. 7.6e-48;
Matches 202; Conservative 55; Mismatches 128; Indels 39; Gaps 8;

QY 54 LPDFAQLVQSEGPVAVNI---QAAPAPRTQNGSNAETSDPLADSDP--FYEFFKRLVP 108
DB 30 LPDFTPLVEQASPAVAVNISTRQKLPDRAVAN-----QQMPDLEGLPMLREFLEERSMP 82
QY 109 NMPEIPQEEADGG-----GLNFGSGFIISKNGYILTTTHVVAGMGSIKVLNDRKREYTA 163
DB 83 PGSRPPGSGAGKGDROREAOGLSGSGFIISPDGYVLTNNHVIDGADELVLRLSDRSELKAK 142
QY 164 LIGSDVQSDVALLKIDATEELPVVKIGNPKNKPGEWVAAGAPFGFDSNVTAGIVSAK 223
DB 143 LVGTDPRTDVAVLKIEG-KDLPTAKLGNSNTLVKGEWVLAIGSPFGFDSHVSVTAGIVSAK 201
QY 224 RSLPNESYTPFTQTDVAINPGNSGGPLNKGQVVGINSQIYRSRSGFGMGISFAIPIDVA 283
DB 202 RSLPNYVPFIQTQTDVAINPGNSGGPLNKGQVVGINSQIYRSRSGFGMGISFAIPIDVA 261
QY 284 MNVAEOLKNTGKVGQGLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343
DB 262 MDVANOLKANGKVSRLGVLGVVIOEVNKLDAESFGLDKPAGLVAQLVEDGPAAGKGVGVG 321
QY 344 DIVLSLDGGEIRSSGDLPMVVGVAITPGKEVSLGVWRKGEETITKAKLGNAAEH----- 396
DB 322 DVILSANGQPIVMSADLPHLIGNLKDGSKAELEVIRDKRKQKLTVTVTGALPDEQEGMDV 381
QY 397 --TGASSKTDEAPYTEQQSGTFSVESAGITLQTHDSSGKHLVVVR--VSDAERAGLRH 452
DB 382 GGTGA-----ERSSNRLGVSVIELTAEQKKSLLDKGGVAIKEVTGPGASLIGLQA 431
QY 453 GDEI 456
DB 432 GDVI 435

RESULT 5
Q9PBA3 PRELIMINARY; PRT; 514 AA.
AC Q9PBA3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PERIPLASMIC PROTEASE.
DE XF2241.
```

OS Xylella fastidiosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
OX Xylella.  
NCBI\_TaxID=2371;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=9A5C;  
RX MEDLINE=20365717; PubMed=10910347;  
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio C.S.,  
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
RA Buenos M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
RA Buato M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,  
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
RA Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,  
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,  
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,  
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
RA Krieger J.E., Kuramie E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,  
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,  
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.F.S.,  
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,  
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,  
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,  
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,  
RA Vagla H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,  
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;  
RT "The genome sequence of the plant pathogen Xylella fastidiosa";  
RL Nature 406:151-159(2000).  
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
DR EMBL: AF004037; AF85040.1; -;  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR00126; Ser\_proteas-V8.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00595; PDZ; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
DR PRINTS: PR00839; V8PROTEASE.  
DR SMART: SM00228; PDZ; 2.  
DR PROSITE: PS0106; PDZ; 2.  
DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
DR Complete proteome; Hydrolase; Serine protease.  
KW Complete proteome.  
SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82BE4C CRC64;

Query Match 36.7%; Score 862.5; DB 16; Length 514;  
Best Local Similarity 46.2%; Pred. No. 8.2e-46;  
Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;

Qy 54 LPDFAQLVSEGPVNVNQAPRTQNGSGNAETSDPLADSPFFYEKKRL-----V 107  
Db LPDFTQLVQVGGVNVNIETVIRKKVGRGRIPLND-----IPEFFRRFGPDFQM 106

Qy 108 PNMPEIQEADDGGL--NFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYAKL 164  
Db PNMPEIQEADDGGL--NFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYAKL 164

Qy 165 IGSVDQSDVALLKIDATEELPVYKIGNPKNLKPGEWAAAGAPGFDNSVTAGIVSAKR 224  
Db IGSVDQSDVALLKIDATEELPVYKIGNPKNLKPGEWAAAGAPGFDNSVTAGIVSAKR 224

Qy 225 SLPNES-YTPFTQTDVAINPGNSGGFLNLKGVGVGINSQIYRSKSGFMGISFAIPDVA 283  
Db SLPNES-YTPFTQTDVAINPGNSGGFLNLKGVGVGINSQIYRSKSGFMGISFAIPDVA 283

Qy 224 STSDQRYVPFIQTDVPINQNSGGFLNTRGEVIGINSQIFSAISGSGYMGISFAIPINLA 283  
Db STSDQRYVPFIQTDVPINQNSGGFLNTRGEVIGINSQIFSAISGSGYMGISFAIPINLA 283

Qy 284 MNVAEOLKNTGKVQRGOLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343  
Db MNVAEOLKNTGKVQRGOLGVIIQEVSYGLAQSGFLDKASGALIAKILPGSPAERAGLQAG 343

Qy 344 DIVLSLDGGEIRSSGDLPMVVGAIPTGKEVSLGWKRGKE--EITIKAKLGNAAHTGASS 401  
Db DIVLSLDGGEIRSSGDLPMVVGAIPTGKEVSLGWKRGKE--EITIKAKLGNAAHTGASS 401

Qy 402 KTDPEIYEQOQGTFSVAGITLOTHDSSGKHLVVVRYSDAAERAGLRHGDILAVRA 461  
Db KTDPEIYEQOQGTFSVAGITLOTHDSSGKHLVVVRYSDAAERAGLRHGDILAVRA 461

Qy 403 EDDTAPSKPETA--NVELLGLQVENLSAERERL-----ASSONAKGGVRITAVTA 452  
Db EDDTAPSKPETA--NVELLGLQVENLSAERERL-----ASSONAKGGVRITAVTA 452

RESULT 6  
Q98CS8 PRELIMINARY; PRT; 513 AA.  
ID Q98CS8  
AC Q98CS8  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE SERINE PROTEASE, HTRA/DEQ/DEGS FAMILY.  
GN MLI5022.  
OS Rhizobium loti (Mesorhizobium loti)  
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
OC Phyllobacteriaceae; Mesorhizobium.  
OX NCBI\_TaxID=381;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=MAFF303099;  
RC MEDLINE=21082930; PubMed=11214968;  
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
RA Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
RT Mesorhizobium loti";  
RL DNA Res. 7:331-338(2000).  
DR EMBL: AP003005; BAB51543.1; -;  
DR MEROPS: S01.UPC; -;  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; Trypsin.  
DR Pfam: PF00595; PDZ; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
DR SMART: SM00228; PDZ; 2.  
DR PROSITE: PS0106; PDZ; 2.  
DR PROSITE: PS0240; TRYPSIN\_DOM; 1.  
KW Complete proteome.  
SQ SEQUENCE 513 AA; 52258 MW; 901FAF72004F5E3B CRC64;

Query Match 31.7%; Score 746; DB 16; Length 513;  
Best Local Similarity 39.0%; Pred. No. 1.5e-38;  
Matches 184; Conservative 81; Mismatches 177; Indels 30; Gaps 11;

Qy 10 LAALCAALLAGCEKAGSFFGADKKKEASFYERIEHTKDDGVSVMLLPD--FAQLVQSEGPA 67  
Db LAALCAALLAGCEKAGSFFGADKKKEASFYERIEHTKDDGVSVMLLPD--FAQLVQSEGPA 67

Qy 68 VVNIQAAAPRTQNGSGNAETSDPLADSPFFYEKKRLVPMN---PEIQEADDGGL 123  
Db VVNIQAAAPRTQNGSGNAETSDPLADSPFFYEKKRLVPMN---PEIQEADDGGL 123

Qy 124 NFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYAKLIGSDVQSDVALLKIDATEE 183  
Db NFGSGFIISKNGYILTNTHVVGMSIKVLLNDKREYAKLIGSDVQSDVALLKIDATEE 183

Qy 124 -LGSQFIVTADGTVTNHNVVDGASSIKVTLDDGTELPKLVGRDAKNDLAVLKISDKP 182  
Db -LGSQFIVTADGTVTNHNVVDGASSIKVTLDDGTELPKLVGRDAKNDLAVLKISDKP 182

Qy 184 LPVVKIGNPKNLKPGEWAAAGAPGFDNSVTAGIVSAKGRSLPNESYTPFIQTDVAINP 243  
Db LPVVKIGNPKNLKPGEWAAAGAPGFDNSVTAGIVSAKGRSLPNESYTPFIQTDVAINP 243

Qy 183 LPTVKWGSDDLMTGDQVLAIGNPFCIGTIVTAGIVSARGDLHSGPFDFFIQIDAPINH 242  
Db LPTVKWGSDDLMTGDQVLAIGNPFCIGTIVTAGIVSARGDLHSGPFDFFIQIDAPINH 242

Qy 244 GNSGGFLNLKGVGVGINSQIYRSKSGFMGISFAIPIDVAMNVAEOLKNTGKVQRGOLCV 303  
Db GNSGGFLNLKGVGVGINSQIYRSKSGFMGISFAIPIDVAMNVAEOLKNTGKVQRGOLCV 303



Db	258	ADQLKATGCKVARGWLGVIIQEVNKNDLAESFGLDRPAGALVAQVLEDPADKGLQVGDVI	317
Qy	347	LSLDGGGIRSGDLPVMVGATTPCKEYVSLGWRRKGEE- - - -ITIKA- - -KLGNAAEHTGAS	400
Db	318	LSLGGHPIVMSADLPHLVGGGLKPGAAANLEWVRDGRKRRTAITVGALEPGNGVQPSIAG	377
Qy	401	SKTDEAPYTEQQSGTFSFESAGITLQTHTDSGKHLVVR- -VSDAAERAGLRHGDILA	458
Db	378	-----TEGSSNRLGVTVTTLTAEQKSLDLKGGVVIREVLNGPAALIGLRPGDVVTH	429
Qy	459	VRASP	463
Db	430	LNNQP	434

```

RESULT      8
FN Q985F9
ID Q985F9 PRELIMINARY; PRT; 516 AA.
AC
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation, update)
DE SERINE PROTEASE.
GN MLR7692.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RT SEQUENCE FROM N.A.
RT STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kashiwa Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003012; BAB54103.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR000126; Ser_proteas_v8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
DR ProSease; Complete proteome.
SQ SEQUENCE 516 AA; 53704 MW;
FE8DAADC099ABDC6 CRC64;

```

```

DR SMARI; SMO0228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Protease; Complete proteome.
SQ SEQUENCE 516 AA; 53704 MW; FE8DAADC099ABDC6 CRC64;

Query Match      30.68;  Score 719.5;  DB 16;  Length 516;
Best Local Similarity 36.28;  Pred. No. 6.6e-37;
Matches 174;  Conservative 80;  Mismatches 177;  Indels 49;  Gaps

Qy 10 LAALCAALLAGCEKAGSPFGADKKEASVERIEHTKDDGSVSMLLPDPFAQLVQSGEPAVV 69
   ||| : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 16 LAAASVAVAGVIGVGALTSGTSPVLADAVRVEAPQVG-----FADVVVERVSPAVV 67

Qy 70 N--IQAAPAPRTQNGNAETSDPLADSDPFYEFFKRLVPNMPEIPOEADDGLN--- 124
   : : : | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 68 SVKVAKAIQPTADGSD-D-QGDFDNLPNPQLRPFKFEF-----RGFGDQGQONDEG 118

Qy 125 -----FGSGCFIISKNGYILTNTHVVVAGMGSIKVLLNDRKREYTKALIG 166
   |||| |||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 119 HRRGHRDRSDNQPRPVAQGGSEFISEDGYLVTNNHVVVEGTAFVTNNMDGKELDALKV 178

Qy 167 SDVQSDVALLKIDATEELPVVYKIGNPKNLKPGEWVAAICAPFGFDNSVTAGIVSAKGRSL 226

```

```

Query Match      28.6%; Score 673; DB 16; Length 504;
Best Local Similarity 36.4%; Pred. No. 5e-34;
Matches 178; Conservative 73; Mismatches 170; Indels 68; Gaps 15;

QY 10 LAALCAALLAGCEKAGSF----FGADKKEASFVERIEHTKDDGVSMLLPDFAQLVQSEG 65
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 14 LIAGTAALLVGTVAAPSFVTPVFAA-----DGPASV-----ADLAQGVL 52

QY 56 PAVVNIQAAPRQTQNGSGNAETSDSLADSDPYEFYEFKRLVPMNPQEAADDGG--- 122
   | | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 53 GAVNNISTQTVKTEGPG--AVPMPLQPGSPQGFDDFFFKNR---GDGKNISGKV 106

```

[illegible]

14	AAVALSASL-----AGAFVVTGPIGLNNEARAE-----AVHVTPPPQAGFADLVEKVR	61
66	PAVVNTQAAPAPRTONGSGN---AETSDPLADSDPYEFF-----KRLV	107
62	PAVSVRVKDVQETSNRGPOFFPGFDQLPDGHPILKREFRDFGMEPRGDSRSDNRKKG	121
108	PNMPEIQEADGGUNFGSGFIISKNGYILTNTHVVVAGMGSIKVLNLDKREYTKALIGS	167
122	ANKPRFCHERP-----VAQSGSEVISEDGYVNNHVVSDGATVVLDDGTDLAKLIGA	177
168	DVQSDVALKIDATELPPVK-IGNPNKLKPCGEWAAGAPPFEDNSVTAGIVSAKGRSL	226
178	DPRDLAVLKINAPKRFVYSPLATIIRCDVGDWVAVGNPFLGCTGTTSGIVSARGADI	237
227	PNESYTPFIOTDVAINPGNSGGPLFNLKGVVGINSOIYSRSGFMGISFAIPIDVAMNV	286
238	GAGPYDDFIQIDAANVKNSGGPAFDLGSVEVIGINTALFSPSGGTGVIAGFAIPSSAKOV	297
287	AEOLKNTGKVQSGOLGVIIQIYESYGLAQSFGLDKASGALITAKILPGSPAERAGLQAGDIV	346
298	VDOLIKKSGVERGWIGVQTPYTKDIAASLGLAEKGAIVASPQDDGPAAKAGIKAGDIV	357
347	LSLDGGEITRSGDPLVMVGAITPGKEVSLGVWR--KGEITIK-AKLGNAEHTGASSKT	403
358	TAVNGETVODPRDLARKVANIAPGEKAALTVWRKNKAEIINVTAAAMPNDKGGSGQSND	417
404	DEAPYTEOQSCTFSVESAGITLTQHTDSSGKHLVVVR---SDAAERAGLRHGDILAV	459
418	NDGGOG-----TLDSTGLTVVPSDE--GKGVVTVDDPDSDAADR-GIRSGDIVSV	467
RESULT	11	
Q9FD11	PRELIMINARY;	PRT; 453 AA.
AC	Q9FD11	
ID	Q9FD11	
DT	01-WAR-2001 (TREMBlrel. 16, Created)	
DT	01-WAR-2001 (TREMBlrel. 16, Last sequence update)	
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)	
DE	HTRA-LIKE SERINE PROTEASE.	
DN	PTSL.	
OS	Aeromonas hydrophila.	
OC	Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;	
OC	Aeromonas.	
OX	NCBI_TaxID:644;	
RN	[1]	
RC	SEQUENCE FROM N.A.	
RC	STRAIN-CKH-29;	
RA	Lin T.-N., Lin T.-J., Liou C.-M.;	
RT	"Aeromonas hydrophila strain CKH-29 prtS1 gene complete cds.";	
RL	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.	
RR	EMBL; AF293977; AA003073.1; -.	
RR	MEROPS; S01.274; -.	
DR	InterPro; IPR001478; PDZ.	
DR	InterPro; IPR001940; Protease2C.	
DR	InterPro; IPR001254; Trypsin.	
DR	Pfam; PF00595; PDZ; 2.	
DR	Pfam; PF00089; trypsin; 1.	
DR	PRINTS; PR00834; PROTEASES2C.	
DR	SMART; SM00228; PDZ; 2.	
DR	PROSITE; PS0106; PDZ; 2.	
DR	PROSITE; PS02040; TRYPSIN_DOM; 1.	
DR	Hydrolase; Protease; Serine protease.	
SW	SEQUENCE 453 AA; 47828 MW; BD55CE2B84E5148 CRC64;	

```

Query Match      28.0%  Score 660:  DB 2:  Length 453;
Best Local Similarity 39.2%:  Pred.No. 2.8e+33;
Matches 160:  Conservative 76:  Mismatches 140:  Indels 32:  Gaps 8:
QY      54 LPDFAQLVQSEGPVAVNTQAAAPAPRTQNGSNAETSDPLADSDPYEFYFKRLVPMKPII 113
      :  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
db      38 MPSLAPVLEQVTPVAVNTVSGKKITQRQL-----PEQFFRFGF---GNMPD- 81

```

QY	114	POEADDDGLNFGSGFTII-SKNGYILTNTHVWAGMSIKVLLNDRKEYTAKLIGSDVQSD	172
Db	82	-EQVSEFPQALGSGVLIIDAKKGYVITNAHVHDEAIKVLNLDKRGYAAKKIGEDKQSD	140
QY	173	VALLKIDATEELPWVIGNPKNLKPEWVAATACAPGFQNSVTAGIVSAKGRS-LPNEYS	231
Db	141	IALLQIKA-EDLVQIKFADSEDLRVGDYALAIQNPGLGQVITVSGIVSALCRSGLNIENL	199
QY	232	TPFTQTDV/ALNPGNSGGPLFNLAGOVVYINSQIYSRSGPGMGISFAIPIDVAMNVAEOLK	291
Db	200	ENF/OTDAALNSGSGGALLNLRGELIGINTAILPGNGNIGIGFAIPSNNVRLDSQIV	259
QY	292	NTGKVRQGLGVIIIOEVSYGUAQSGFLDKASGALIAKILPCSPAERAGLAGODIVLSLDG	351
Db	260	KYGEVRRQGLGIICTELTSEVAKTFEGYKKDGAFVQVMPDSPAANKAGIKPGDIIVSIDG	319
QY	352	GEIRSSGDLPMVGAITPGKEVSLGVWRKGBEITIKARLGNAAEHTGASKTSDPAPYTEQ	411
Db	320	KAIRSFGELRAKIIATMGADKQVAILGIFDQKEQIVKVTLLKADDSEILASALHPA----	374
QY	412	QSGTFSVESAGITIQTHTDSSGKHLVVRVSDAAERACLRHGDEILAV	459
Db	375	-----LECAKLS-TTSEPVSGVAVSEIDPRSPAASGLQKGDVIIGV	415
RESULT	12		
ID	P74978	PRELIMINARY:	PRT: 478 AA.
AC	P74978;		
DC	01-FEB-1997 (TREMBLrel. 02, Created)		
DT	01-FEB-1997 (TREMBLrel. 02, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	GSRA PROTEIN.		
GN	GSRA.		
OS	Yersinia enterocolitica.		
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		
OC	Yersinia.		
OX	NCBI_TaxID=630;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-WA314;		
RX	MEDLINE=96333330; PubMed=8757824;		
RA	Yamamoto T., Hanawa T., Ogata S., Kamiya S.;		
RT	"Identification and characterization of the Yersinia enterocolitica		
RT	gsrA gene, which protectively responds to intracellular stress induced		
RT	by macrophage phagocytosis and to extracellular environmental stress.;"		
RL	Infect. Immun. 64:2980-2987(1996).		
DR	EMBL, D78376; BAA11382.1; -		
DR	MEROPS; S01.273; -		
DR	InterPro; IPR001478; PDZ.		
DR	InterPro; IPR001940; Protease2C.		
DR	InterPro; IPR001254; Trypsin.		
DR	Pfam; PF00595; PDZ; 2.		
DR	Pfam; PF00089; trypsin; 1.		
DR	PRINTS; PR00834; PROTEASES2C.		
DR	SMART; SM00228; PDZ; 2.		
DR	PROSITE; PS50106; PDZ; 2.		
DR	PROSITE; PS50240; TRIPSIN_DOM; 1.		
DR	Hydrolase; Serine protease.		
SW	SEQUENCE 478 AA; 49459 MW; EDD43CDAL8BDE17C CRC64;		

Query Match	27.2%;	Score 639.5;	DB 2;	Length 478;
Best Local Similarity	35.3%;	Pred. NO. 5.6e-32;		
Matches 154; Conservative		79; Mismatches 162;	Indels 41;	Gaps 8;

[illegible]

	Qy	291	KNTCKVQRGOLGVIIQEIVSYGLAQSFCLDKASGALIAKILPGSPAERAGLQAGDIWLSLD	350
			: : : : :           : : : : :         :	
	Db	290	KKDCKVSRRGLGVYIIDLTEIDISEGLCLKNTRGVLAQVKVEDPGDKAGIKTGDIIIEFA	349
			: : : : :           : : : : :         :	
	Qy	351	GGETRSSGDPLVMVGAIITPKCEVSLGWVRKGEEITTKALCNAEHT--GASSCTDEAPY	408
			: : : : :         : : : : :         : : : : :         :	
	Db	350	DIPVNKTKLRLVIIADAPIQDEVKKVLTRDKLELPKIKITSNEEVTKDSTEETNKKEI	409
			: : : : :         : : : : :         : : : : :         :	
	Qy	409	TEQSQTFSVESAGITL-----QTHDTSSCKHLVVVVSVDAARAGLRPHGDEI	456
			: : : : :         : : : : :         : : : : :         :	
	Db	410	TNKENNLSTTKNNITFGNLTEELRORKYTIPODKMGVIITNID-EESSPFKIGDLI	464
			: : : : :         : : : : :         : : : : :         :	

RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NTHI 12;
RX	MEDLINE=98147697; PubMed=9488373;
RA	Loosmore S.M., Yang Y.P., Oomen R., Shortreed J.M., Coleman D.C.,
RT	Klein M.H.;
RR	"The Haemophilus influenzae HtrA protein is a protective antigen.";
RL	Infect. Immun. 66:899-906(1998).
DR	EMBL: AF018151; AAC38202.1; ..
MR	MEROPS: S01_274; ..
DR	InterPro: IPR001314; Chymotrypsin.
DR	InterPro: IPR001478; PDZ.
DR	InterPro: IPR001940; Protease2C.
DR	InterPro: IPR001254; Trypsin.
DR	Pfam: PF00595; PDZ; 2.
DR	Pfam: PF00089; trypsin; 1.
DR	PRINTS: PR00722; CHYMOTRYPSIN.
DR	PRINTS: PR00834; PROTEASES2C.
DR	SMART: SM00228; PDZ; 2.
DR	PROSITE: PS50106; PDZ; 2.
DR	PROSITE: PS50240; TRYPSIN_DOM; 1.
KW	Hydrolase; Serine protease.
FT	1
SQ	SEQUENCE   437 AA;  46376 MW;   27AE2E9F27BE5F6C CRC64;

Query Match	27.1%	Score	637;	DB	2;	Length	437;
Best Local Similarity	38.7%;	Pred.	No. 7.le-32;				
Matches	161;	Conservative	75;	Mismatches	152;	Indels	28;
Gaps	12;						
Qy	54	LPDFAQLVQSGP	PAVVNIQAAPRTONGSCNAETDS-DPLADSDP--FYEFF-KRLVYN	109			
Db							
Db	2	LPSFVSEONS	UAPMLEKVQ--PAVVTVSVEGAKVDSSPFLDDIPEDKEFFGURFAEQ	59			
Qy	110	MPEIQEAD	DGGLNFGSGFTI-SKNGYILNTHTHVACMGSIKVLNLDKREYTKAKIGSD	168			
Db							
Db	60	FGGRGESKRN	FRGL--CSGVIINASKGVYLTNNHVIDGADKITVLOQDGRFEKAKLVCKD	117			
Qy	169	VQSDVALLK	IDATELPVVKTCNPKNLKPGEWAAIAGAPFGFDNSVTAGIVSAGKRSLPN	228			
Db							
Db	118	EQSDIALVQLEK	PNLTKIETKFDADSKLRVGDVTVAIGNPFLGTQVTTSGIVSALGRSTGS	177			
Qy	229	ES--YTPFIQTD	VAIIPNGNSGGLFNLKQGVVGNISQIYSRSGGFMGTSFAIPIDVANNV	286			
Db							
Db	178	DSGTIYENVIQ	TDAAVNRGNSGCGALVNLGELIGINTAILSPSGGNAGIATAFIPSNQASNL	237			
Qy	287	AEQLKNTGK	VQVRGVIIEQVSYSGAQSFGLDKASGALIAKIILPGSPAERAGLOACDIV	346			
Db							

Db 313 LKAGDVIVAMNGOKISSFAEMRAKATATSGAGKEIALTYLRDCK-----THOTKV 361  
QY 400 SSKTDEAPYTEOOSGTFVSAGITLQHTDSSGKHLVVVRVS--DAAERAGLRHGDEIL 457  
Db 362 TLOSDD--QTQADASHLLPALAGAEMSNHDKGVKGLITSVTPKSLAEOGLKKGDVII 419  
QY 458 AV 459  
Db 420 GV 421

Search completed: June 10, 2002, 12:13:55  
Job time: 239 sec

Db 238 VQOILEFCQVRGLGIGKGELNADLAKAFNVSAQQAFVSEVLPKSAEAKGLKAGDII 297  
QY 347 LSLDGEIRSSGDLPLVMVGAITPCGEYSLGVWRKGEIITIKAKLGNAAEHTGASSKTTDEA 406  
Db 298 TAMNQKTISSFAEIRAKIATTGAKREISLYLRDCKSHDKMKL-QADDGSQLSSKT-EL 355  
QY 407 PYTEOOSGTFVSAGITLQHTDSSGKHLVVVRV---SDAAERAGLRHGDEILAV 459  
Db 356 PALD-----GATLKDYDAKGVKGIEITKIQPNSLAAQR-GLKSGDIIIGI 399

RESULT 15  
Q9CMS7  
ID Q9CMS7 PRELIMINARY: PRT: 459 AA.  
AC Q9CMS7:  
DT 01-JUN-2001 (TremBLrel. 17, Created)  
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)  
DE HTRA.  
GN HTRA OR PM0734.  
SS Pasteurella multocida.  
S Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
CC Pasteurella.  
OX NCBI\_TaxID=747;  
RN [1]  
SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RX MEDLINE=21145866; PubMed=11248100;  
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida Pm70.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).  
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.  
DR EMBL: AE006110; AAK02818.1;  
DR MEROPS: S01\_274;  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001940; Protease2C.  
DR InterPro: IPR001254; trypsin.  
DR Pfam: PF00595; PDZ; 2.  
DR Pfam: PF00089; trypsin; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
DR SMART: SM00228; PDZ; 2.  
DR SMART: SM00020; Tryp\_Spc; 1.  
DR PROSITE: PS50106; PDZ; 2.  
DR PROSITE: PS50240; TRYPSIN\_DOM; 1.  
DR Complete proteome; Hydrolase; Serine protease.  
KW Complete proteome; Hydrolase; Serine protease.  
SQ SEQUENCE 459 AA: 48533 MW: 9889E3945812F0FA CRC64;  
  
Query Match 27.1%; Score 637; DB 16; Length 459;  
Best Local Similarity 35.8%; Pred. No. 7.6e-32;  
Matches 151; Conservative 85; Mismatches 150; Indels 36; Gaps 9;  
QY 44 TKDDGSVMLLPDPFAQLVQSEGPAVVNIQAAAPRTQNGSNAET-DSDPLADSDPFYEF 102  
Db 30 TVDVG--QOVPSLAPMLEKVLPAVVSV-----EGKAKNHGSTPFQGIPEPRF 77  
QY 103 EKRLVPMPEIPOEADGGLNFGSGFII-SKNGYILTNTHTVVVAGMGSTKVLINDKREYT 161  
Db 78 FFG-----PDLFNDRAPRNFRTGIGSGVIIINAEKGYVLTNNHVIDSADKITVQLQDGRLS 132  
QY 162 AKLGSDVQSDVALLKIDATEELPVYKIGNPKLKGEEVAAJICAPFGFDSNVTAGIVSA 221  
Db 133 AKVIGTDELSDVALIQIEKPKNLTAJLTDSDKLRVGDFTVAIGNPFLGQTVTSGIVSA 192  
QY 222 KGRSL--PNESYTPFIQTDVAINPGNSGGPLFNLKGQVGINSGIYSRSGFGMGSFAIP 279  
Db 193 LGRSTGSDNGHYENYIQTDAVNRNGSGGPLINLQGLIGINTAILSPSGGNAGIAFAIP 252  
QY 280 IDVAMNVAEQLNKTKVQVGOLGVIQFVSYGLAQSGLDKASGALIAKILPGSPAERAG 339  
Db 253 SNWANNLVQIILEFGEVRRGMLGKGGELNADLAKAFDIEAQQAQAFVSEVLPSNAEAKG 312  
QY 340 LQAGDIVLSLDGGEIRSSGDLPLVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGA 399

